

166053

STIC-Biotech/ChemLib

mg

From: Page, Thurman
Sent: Sunday, September 18, 2005 3:33 PM
To: Swope, Sheridan; Chan, Christina; STIC-Biotech/ChemLib
Cc: Woodward, Michael; Richter, Johann; Low, Christopher; Fredman, Jeffrey; Page, Thurman
Subject: RE: 10/719,236

Importance: High

RUSH SEARCH APPROVED

Paul
Schulwitz

RECEIVED
SEP 19 2005
STIC-BIOTECH/ChemLib
(STIC)

-----Original Message-----

From: Swope, Sheridan
Sent: Sunday, September 18, 2005 12:56 PM
To: Chan, Christina
Cc: Woodward, Michael; Page, Thurman; Richter, Johann; Low, Christopher; Fredman, Jeffrey
Subject: 10/719,236

May I have this rushed for an allowance?
Pls reply to all, so everyone knows the request has been processed.

For 10/719,236, pls search and interference search:

The translation product of SID 1, residues 232-1164, against the NT and AA data bases.

Note: the translation product's N-terminal and C-terminal residues are:
Met-Asp-Thr-Thr-Val-Pro.....Arg-Arg-Thr-Ser-Lys-Ala

Thank you,

Sheridan Swope, Ph.D.
Patent Examiner, AU 1656
Recombinant Enzymes
571-272-0943 (voice)
E02B71 Remsen Bld (Office)
E03C70 Remsen Bld (Mailbox)

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: _____
Date completed: _____
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# _____ AA# _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other (Specify): _____

THIS PAGE BLANK (uspro)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2005, 15:00:07; Search time 54 Seconds

(without alignments)
2227.453 Million cell updates/sec

Title: 10719236-1_232-1164

Perfect score: 1636

Sequence: 1 MDTVPTFSLAELOGGLHOD.....ATFDWIGNVNIRRTSKA 311

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq.16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1636	100.0	311	7	ADE47638
2	1634	99.9	311	2	AAW17775
3	1634	99.9	311	5	AAU11044
4	1632	99.8	311	5	AAU11050
5	1631	99.7	311	5	AAU11046
6	1631	99.7	311	5	AAU11051
7	1630	99.6	310	5	AAU11053
8	1630	99.6	310	5	AAU11525
9	1630	99.6	311	1	AAW34168
10	1630	99.6	311	2	AAW34176
11	1630	99.6	311	2	AAW34176
12	1630	99.6	311	2	AAW34176
13	1629	99.6	311	2	AAW34175
14	1629	99.6	311	4	AAU11016
15	1629	99.6	311	5	AAU11047
16	1629	99.6	311	6	ABG73769
17	1628	99.5	311	2	AAW34167
18	1628	99.5	311	2	AAW34177
19	1628	99.5	311	5	AAU11048
20	1627	99.4	310	5	AAU11528
21	1627	99.4	311	2	AAW22220
22	1627	99.4	311	5	AAU11052
23	1626	99.4	311	2	AAW34169
24	1626	99.4	311	2	AAW34174
25	1626	99.4	311	2	AAW34166

26	1626	99.4	311	5	AAU11049
27	1625	99.3	309	5	AAU11054
28	1625	99.3	309	5	AAU11526
29	1625	99.3	310	5	AAU11535
30	1625	99.3	310	5	AAU11532
31	1625	99.3	310	5	AAU11531
32	1625	99.3	310	5	AAU11534
33	1625	99.3	310	5	AAU11533
34	1625	99.3	311	2	AAW34173
35	1625	99.3	311	2	AAW34171
36	1625	99.3	311	2	AAW40425
37	1624	99.3	311	2	AAW34172
38	1624	99.3	311	2	AAW34170
39	1624	99.3	311	2	AAW40426
40	1623	99.2	311	2	AAW40428
41	1623	99.2	311	2	AAW40428
42	1622	99.1	309	5	AAU11529
43	1621	99.1	308	5	AAU11521
44	1621	99.1	308	5	AAU11527
45	1621	99.1	311	2	AAW40427

ALIGNMENTS

RESULT 1	ADW47638	standard; protein; 311 AA.
XX	ADW47638;	
XX	29-JAN-2004	(first entry)
XX	DE	S. clavuligerus penicillin expandase SEQ ID NO:2.
XX	penicillin expandase; penicillin G; 7-aminodesacetoxycephalosporanic acid.	
XX	Streptomyces clavuligerus.	
XX	EP1348759-A1.	
XX	01-OCT-2003.	
XX	29-JAN-2003; 2003EP-00250544.	
XX	26-MAR-2002; 2002US-00105319.	
XX	(SYNM-) SYNMAX BIOCHEMICAL CO LTD.	
XX	Yang Y, Wei C, Hsu J, Tsai Y;	
XX	WPI; 2003-781116/74.	
XX	N-PSDB; ADE47637.	
XX	New mutated penicillin expandase having an expandase activity towards penicillin G, which is at least at 2-fold higher than the activity of wild-type expandase, useful for producing 7-aminodesacetoxycephalosporanic acid (7-ADCA).	
XX	Disclousure; SEQ ID NO 2; 55pp; English.	
XX	The invention relates to a novel mutated penicillin expandase having an expandase activity towards penicillin G, which is at least at 2-fold higher than the activity of wild-type expandase. The mutated penicillin expandase is useful for producing 7-aminodesacetoxycephalosporanic acid. The present sequence represents the wild-type penicillin expandase.	
XX	Sequence 311 AA;	
XX	Query Match	100.0%; Score 1636; DB 7; Length 311;
XX	Best Local Similarity	100.0%; Pred. No. 6,4e-168;
XX	Matches 311; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 MDITVPTFSLAEIQQGLHODEFRRCRLRDKGLFYLTDGCLTDTELKSAKDVIDFEEHGE 60
Db      1 MDITVPTFSLAEIQQGLHODEFRRCRLRDKGLFYLTDGCLTDTELKSAKDVIDFEEHGE 60
Qy      61 AEKRAVTSVPVPTMRGFTGLESESTAOITNTGYSYSDYSMCYSMGTAADNLFPSGDFERIMT 120
Db      61 AEKRAVTSVPVPTMRGFTGLESESTAOITNTGYSYSDYSMCYSMGTAADNLFPSGDFERIMT 120
Qy      121 QYFDRQYTSRAVARAVLRATGTEPDGVEAFIDCEPLLRFRYPQVPEHRSABEQPLRM 180
Db      121 QYFDRQYTSRAVARAVLRATGTEPDGVEAFIDCEPLLRFRYPQVPEHRSABEQPLRM 180
Qy      181 APHYDLSMTVTLIQOTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVGGQVKA 240
Db      181 APHYDLSMTVTLIQOTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVGGQVKA 240
Qy      241 PRHHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSVPLARECGFDVSLDGETATFQDWIGG 300
Db      241 PRHHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSVPLARECGFDVSLDGETATFQDWIGG 300
Qy      301 NYVNIIRRTSKA 311
Db      301 NYVNIIRRTSKA 311

```

RESULT 2

AAU17775

ID AAU17775 standard; protein; 311 AA.

XX AAU17775;

DT 17-FEB-1998 (first entry)

DE S. clavuligerus wild type expandase.

XX Expandase; 7-aminodesacetoxycephalosporanic acid; 7-ADCA production;
 KM thiazolidine ring expansion; penicillin N; dihydrothiazine ring; enzyme;
 KM semi-synthetic cephalosporin production; penicillin G; isopenicillin N;
 KM cephalosporin ring; DOAC.

XX Streptomyces clavuligerus.

XX MO9720053-A2.

XX 05-JUN-1997.

XX 27-NOV-1996; 96WO-EP005358.

XX 27-NOV-1995; 95EP-00203259.

XX 27-NOV-1995; 95US-0007564P.

XX (KONN) GIST-BROCADES BV.

XX (UYOX-) UNITV OXFORD.

XX Sutherland JD, Bovenberg RAL, Van Der Jaan JM;

XX WPI; 1997-310608/28.

XX Expandase enzymes with mutation(s) to alter substrate specificity - used
 PT particularly for production of 7-amino:desacetoxy:cephalosporanic acid.

XX Disclosure; Fig 1; 21pp; English.

XX This sequence represents the wild type expandase from Streptomyces
 CC clavuligerus. The DNA encoding this sequence was subjected to site
 CC directed mutagenesis to create the mutant expandase sequences of the
 CC invention (see AAU34166-W34177). The mutant expandase sequences have an
 CC altered substrate specificity compared to the wild type expandase.
 CC Expandase catalyses the expansion of the 5-membered thiazolidine ring of
 CC penicillin N to the 6-membered dihydrothiazine ring of DOAC. The mutant
 CC expandase gene is used particularly for the production of 7-
 CC aminodesacetoxycephalosporanic acid (7-ADCA). 7-ADCA is one of the key
 CC intermediates that is used in the preparation of semi-synthetic

CC cephalosporins. The mutant expandases have improved ability to expand
 CC penicillin G into a cephalosporin ring in vitro and/or in vivo where
 CC other penicillins such as penicillin N and isopenicillin N can act as
 CC competing substrates

SQ Sequence 311 AA;

Query Match 99.9%; Score 1634; DB 2; Length 311;
 Best Local Similarity 99.7%; Pred. No. 1.1e-167;
 Matches 310; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 MDITVPTFSLAEIQQGLHODEFRRCRLRDKGLFYLTDGCLTDTELKSAKDVIDFEEHGE 60
Db      1 MDITVPTFSLAEIQQGLHODEFRRCRLRDKGLFYLTDGCLTDTELKSAKDVIDFEEHGE 60
Qy      61 AEKRAVTSVPVPTMRGFTGLESESTAOITNTGYSYSDYSMCYSMGTAADNLFPSGDFERIMT 120
Db      61 AEKRAVTSVPVPTMRGFTGLESESTAOITNTGYSYSDYSMCYSMGTAADNLFPSGDFERIMT 120
Qy      121 QYFDRQYTSRAVARAVLRATGTEPDGVEAFIDCEPLLRFRYPQVPEHRSABEQPLRM 180
Db      121 QYFDRQYTSRAVARAVLRATGTEPDGVEAFIDCEPLLRFRYPQVPEHRSABEQPLRM 180
Qy      181 APHYDLSMTVTLIQOTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVGGQVKA 240
Db      181 APHYDLSMTVTLIQOTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVGGQVKA 240
Qy      241 PRHHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSVPLARECGFDVSLDGETATFQDWIGG 300
Db      241 PRHHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSVPLARECGFDVSLDGETATFQDWIGG 300
Qy      301 NYVNIIRRTSKA 311
Db      301 NYVNIIRRTSKA 311

```

RESULT 3

AAU11044

ID AAU11044 standard; protein; 311 AA.

XX AAU11044;

XX 12-MAR-2002 (first entry)

DE Streptomyces clavuligerus penicillin N expandase protein sequence.

XX Penicillin N expandase; ring-expanding; penicillin G;
 KM phenylacetyl-7-ADCA; amino desacetoxyccephalosporanic acid; cephalosporin;
 KM cephalaxlin; penicillin V.

XX Streptomyces clavuligerus.

XX MO200185951-A1.

XX 15-NOV-2001.

XX 09-MAY-2001; 2001MO-GB002047.

XX 09-MAY-2000; 2000GB-00011185.

XX (ACSD-) ACS DOBPAR UK LTD.

XX Johnson RI, Newbert RW;

XX WPI; 2002-075247/10.

XX N-PADB; AAS17241.

XX New enzyme useful for ring-expanding penicillin G to produce phenylacetyl
 PT -7-amino desacetoxyccephalosporanic acid, comprises penicillin expandase
 PT having increased specificity for substrates such as penicillin G.

XX Claim 4; Fig 1; 42pp; English.

CC The present invention relates to new penicillin expandases modified to
CC improve the ring-expanding activity for a substrate which is not the
CC natural substrate of the unmodified expandase. The invention is useful
CC for ring-expansion of penicillin G to produce phenylacetyl-7-ADCA (amino
CC desacetoxycephalosporanic acid) and the phenylacetyl side chain is
CC removed from phenylacetyl-7-ADCA to produce 7-ADCA, which is useful as a
CC starting point for the production of a range of semi-synthetic
CC cephalosporins, most notably cephalixin. The polynucleotide of the
CC invention is useful for producing a primer e.g. a PCR primer and as a
CC probe. The modification of the penicillin expandase enhances the activity
CC of the enzyme such as penicillin N expandase for penicillin G or V as a
CC substrate. The modified expandase has enhanced catalytic activity or
CC increased specificity for another substrate such as penicillin G. The
CC present protein sequence represents the penicillin N expandase enzyme of
CC the invention

XX SQ Sequence 311 AA;

Query Match 99.9%; Score 1634; DB 5; Length 311;

Best Local Similarity 99.7%; Pred. No. 1.1e-167; Mismatches 0; Gaps 0;

Matches 310; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDTTVPFSLAELOOGLHODEFRRLCKGLFYLTDCGLTDTLKSANDVIDPFEHSGE 60

DB 1 MDTTVPFSLAELOOGLHODEFRRLCKGLFYLTDCGLTDTLKSANDVIDPFEHSGE 60

QY 61 AEKRAVTSVPVPMRGFTGLESESTAOITNTGSDYSVMCGTADNLPFSGDPERIWT 120

DB 61 AEKRAVTSVPVPMRGFTGLESESTAOITNTGSDYSVMCGTADNLPFSGDPERIWT 120

QY 121 QYFDRQYASRAVAREVLARATGTEPDGVEAFDCEPLRFYFPQVPEHRSABEQPIRM 180

DB 121 QYFDRQYASRAVAREVLARATGTEPDGVEAFDCEPLRFYFPQVPEHRSABEQPIRM 180

QY 181 APHYDLSWVTLTIQOTPCANGFVSLQAEVGAFDTLPYRPDAVLVFCGAIATLVGSGQYKA 240

DB 181 APHYDLSWVTLTIQOTPCANGFVSLQAEVGAFDTLPYRPDAVLVFCGAIATLVGSGQYKA 240

QY 241 PRHHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSPVLARECGFVDSLGETATFQDWIGG 300

DB 241 PRHHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSPVLARECGFVDSLGETATFQDWIGG 300

QY 301 NYVNIRRTSKA 311

DB 301 NYVNIRRTSKA 311

RESULT 4

AAU11050 ID AUU11050 standard; protein; 311 AA.

AC AUU11050;

DT 12-MAR-2002 (first entry)

DE Streptomyces clavuligerus penicillin N expandase 1305L mutant.

XX Penicillin N expandase; ring-expanding; penicillin G;

KW phenylacetyl-7-ADCA; amino desacetoxycephalosporanic acid; cephalosporin;

XX cephalixin; penicillin V; mutant; mutein.

OS Streptomyces clavuligerus.

XX Synthetic.

FT Key Location/Qualifiers

FT Misc-difference 305 /note= "Wild-type Ile substituted by Leu"

XX MO200185951-A1.

XX 15-NOV-2001.

XX 09-MAY-2001; 2001WO-GB002047.

XX 09-MAY-2000; 2000GB-00011185.

PA (ACSD-) ACS DOBFAR UK LTD.

PI Johnson RI, Newbert RW;

DR WPI, 2002-075247/10.

PT New enzyme useful for ring-expanding penicillin G to produce phenylacetyl
PT -7-amino desacetoxycephalosporanic acid, comprises penicillin expandase
PT having increased specificity for substrates such as penicillin G.

PS Claim 8; Page: 42pp; English.

CC The present invention relates to new penicillin expandases modified to
CC improve the ring-expanding activity for a substrate which is not the
CC natural substrate of the unmodified expandase. The invention is useful
CC for ring-expansion of penicillin G to produce phenylacetyl-7-ADCA (amino
CC desacetoxycephalosporanic acid) and the phenylacetyl side chain is
CC removed from phenylacetyl-7-ADCA to produce 7-ADCA, which is useful as a
CC starting point for the production of a range of semi-synthetic
CC cephalosporins, most notably cephalixin. The polynucleotide of the
CC invention is useful for producing a primer e.g. a PCR primer and as a
CC probe. The modification of the penicillin expandase enhances the activity
CC of the enzyme such as penicillin N expandase for penicillin G or V as a
CC substrate. The modified expandase has enhanced catalytic activity or
CC increased specificity for another substrate such as penicillin G. This
CC sequence is one of a collection of Streptomyces clavuligerus mutants,
CC AU11046-AU11054 and AU11521-AU11537, used in the invention to improve
CC the ring-expanding activity of expandase. Note: The present protein
CC sequence is not shown in the specification but is derived from the
CC Streptomyces clavuligerus wild-type penicillin N expandase sequence given
CC in Figure 1 (see AU11044)

XX SQ Sequence 311 AA;

Query Match 99.8%; Score 1632; DB 5; Length 311;

Best Local Similarity 99.4%; Pred. No. 1.7e-167; Mismatches 0; Gaps 0;

Matches 309; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDTTVPFSLAELOOGLHODEFRRLCKGLFYLTDCGLTDTLKSANDVIDPFEHSGE 60

DB 1 MDTTVPFSLAELOOGLHODEFRRLCKGLFYLTDCGLTDTLKSANDVIDPFEHSGE 60

QY 61 AEKRAVTSVPVPMRGFTGLESESTAOITNTGSDYSVMCGTADNLPFSGDPERIWT 120

DB 61 AEKRAVTSVPVPMRGFTGLESESTAOITNTGSDYSVMCGTADNLPFSGDPERIWT 120

QY 121 QYFDRQYASRAVAREVLARATGTEPDGVEAFDCEPLRFYFPQVPEHRSABEQPIRM 180

DB 121 QYFDRQYASRAVAREVLARATGTEPDGVEAFDCEPLRFYFPQVPEHRSABEQPIRM 180

QY 181 APHYDLSWVTLTIQOTPCANGFVSLQAEVGAFDTLPYRPDAVLVFCGAIATLVGSGQYKA 240

DB 181 APHYDLSWVTLTIQOTPCANGFVSLQAEVGAFDTLPYRPDAVLVFCGAIATLVGSGQYKA 240

QY 241 PRHHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSPVLARECGFVDSLGETATFQDWIGG 300

DB 241 PRHHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSPVLARECGFVDSLGETATFQDWIGG 300

QY 301 NYVNIRRTSKA 311

DB 301 NYVNIRRTSKA 311

RESULT 5

AAU11046 ID AUU11046 standard; protein; 311 AA.

AC AUU11046;

DT 12-MAR-2002 (first entry)

```

XX  Streptomyces clavuligerus penicillin N expandase 1156V mutant.
DE
XX
XX  Penicillin N expandase; ring-expanding; penicillin G;
KM  phenylacetyl-7-ADCA; amino desacetoxycephalosporanic acid; cephalosporin;
KW  cephalalexin; penicillin V; mutant; mutein.
XX
XX  Streptomyces clavuligerus.
OS  Synthetic.
XX
XX  Key Location/Qualifiers
FH  Misc-difference 158
FT  /note= "Wild-type Leu substituted by Val"
XX
XX  W0200185951-A1.
PN
XX  15-NOV-2001.
XX
XX  09-MAY-2001; 2001WO-GB002047.
XX
XX  09-MAY-2000; 2000GB-00011185.
XX
XX  (ACSD-) ACS DOBPAR UK LTD.
PA
XX  Johnson RI, Newbert RW;
PI
XX  WPI; 2002-075247/10.
DR
XX
XX  New enzyme useful for ring-expanding penicillin G to produce phenylacetyl
PT  -7-amino desacetoxycephalosporanic acid, comprises penicillin expandase
PT  having increased specificity for substrates such as penicillin G.
XX
XX  Claim 6; Page; 42pp; English.
PS
XX
XX  The present invention relates to new penicillin expandases modified to
CC  improve the ring-expanding activity for a substrate which is not the
CC  natural substrate of the unmodified expandase. The invention is useful
CC  for ring-expansion of penicillin G to produce phenylacetyl-7-ADCA (amino
CC  desacetoxycephalosporanic acid) and the phenylacetyl side chain is
CC  removed from phenylacetyl-7-ADCA to produce 7-ADCA, which is useful as a
CC  starting point for the production of a range of semi-synthetic
CC  cephalosporins, most notably cephalalexin. The polynucleotide of the
CC  invention is useful for producing a primer e.g. a PCR primer and as a
CC  probe. The modification of the penicillin expandase enhances the activity
CC  of the enzyme such as penicillin N expandase for penicillin G or V as a
CC  substrate. The modified expandase has enhanced catalytic activity or
CC  increased specificity for another substrate such as penicillin G. This
CC  sequence is one of a collection of Streptomyces clavuligerus mutants,
CC  AAU11046-AAU11054 and AAU11521-AAU11537, used in the invention to improve
CC  the ring-expanding activity of expandase. Note: The present protein
CC  sequence is not shown in the specification but is derived from the
CC  Streptomyces clavuligerus wild-type penicillin N expandase sequence given
CC  in Figure 1 (see AAU11044)
XX
XX  Sequence 311 AA;
SQ
XX
XX  Query March 99.7%; Score 1631; DB 5; Length 311;
XX  Best Local Similarity 99.4%; Pred. No. 2.2e-167;
XX  Matches 309; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
XX  1 MDTTVPFTSLAELOOGLHODEFFRRLNDKGLFYLTDCGLDTEKSAKDIYIDFFEHGSE 60
DB 1 MDTTVPFTSLAELOOGLHODEFFRRLNDKGLFYLTDCGLDTEKSAKDIYIDFFEHGSE 60
XX
XX  61 AEKRAVTSVPVTRMGRFTGLESESTAOITNTGYSYSDYSCYSGMTADNLFPSGDPERIWT 120
QY 61 AEKRAVTSVPVTRMGRFTGLESESTAOITNTGYSYSDYSCYSGMTADNLFPSGDPERIWT 120
DB 61 AEKRAVTSVPVTRMGRFTGLESESTAOITNTGYSYSDYSCYSGMTADNLFPSGDPERIWT 120
XX
XX  121 QYPRORATASAAVAREVLRATGTBPDGVEAFLLCEPLTARRYPPOVPEHRSABEQPLRM 180
QY 121 QYPRORATASAAVAREVLRATGTBPDGVEAFLLCEPLTARRYPPOVPEHRSABEQPLRM 180
DB 121 QYPRORATASAAVAREVLRATGTBPDGVEAFLLCEPLTARRYPPOVPEHRSABEQPLRM 180
XX
XX  181 APHYDLSMTVTLIOOTPCANGFVSLQAEVGGAFTDLPYRPAVALVFCGAIATLVTGQVKA 240
QY 181 APHYDLSMTVTLIOOTPCANGFVSLQAEVGGAFTDLPYRPAVALVFCGAIATLVTGQVKA 240

```

```

DB 181 APHYDLSMTVTLIOOTPCANGFVSLQAEVGGAFTDLPYRPAVALVFCGAIATLVTGQVKA 240
XX
XX  |||||
QY 241 PRHHVAAPRRDQIAGSSRTSSVFFLEPNADFTFSVFLAECGFDLDETFATFOWMIG 300
DB 241 PRHHVAAPRRDQIAGSSRTSSVFFLEPNADFTFSVFLAECGFDLDETFATFOWMIG 300
XX
XX  301 NYNIRRTSKA 311
QY 301 NYNIRRTSKA 311
DB 301 NYNIRRTSKA 311
XX
XX  RESULT 6
XX  AAU11051
XX  ID AAU11051 standard; protein; 311 AA.
XX
XX  AAU11051;
AC
XX
XX  12-MAR-2002 (first entry)
DT
XX
XX  Streptomyces clavuligerus penicillin N expandase 1305M mutant.
DE
XX
XX  Penicillin N expandase; ring-expanding; penicillin G;
KM  phenylacetyl-7-ADCA; amino desacetoxycephalosporanic acid; cephalosporin;
KW  cephalalexin; penicillin V; mutant; mutein.
XX
XX  Streptomyces clavuligerus.
OS  Synthetic.
XX
XX  Key Location/Qualifiers
FH  Misc-difference 305
FT  /note= "Wild-type Ile substituted by Met"
XX
XX  W0200185951-A1.
PN
XX  15-NOV-2001.
XX
XX  09-MAY-2001; 2001WO-GB002047.
XX
XX  09-MAY-2000; 2000GB-00011185.
XX
XX  (ACSD-) ACS DOBPAR UK LTD.
PA
XX  Johnson RI, Newbert RW;
PI
XX  WPI; 2002-075247/10.
DR
XX
XX  New enzyme useful for ring-expanding penicillin G to produce phenylacetyl
PT  -7-amino desacetoxycephalosporanic acid, comprises penicillin expandase
PT  having increased specificity for substrates such as penicillin G.
XX
XX  Claim 8; Page; 42pp; English.
PS
XX
XX  The present invention relates to new penicillin expandases modified to
CC  improve the ring-expanding activity for a substrate which is not the
CC  natural substrate of the unmodified expandase. The invention is useful
CC  for ring-expansion of penicillin G to produce phenylacetyl-7-ADCA (amino
CC  desacetoxycephalosporanic acid) and the phenylacetyl side chain is
CC  removed from phenylacetyl-7-ADCA to produce 7-ADCA, which is useful as a
CC  starting point for the production of a range of semi-synthetic
CC  cephalosporins, most notably cephalalexin. The polynucleotide of the
CC  invention is useful for producing a primer e.g. a PCR primer and as a
CC  probe. The modification of the penicillin expandase enhances the activity
CC  of the enzyme such as penicillin N expandase for penicillin G or V as a
CC  substrate. The modified expandase has enhanced catalytic activity or
CC  increased specificity for another substrate such as penicillin G. This
CC  sequence is one of a collection of Streptomyces clavuligerus mutants,
CC  AAU11046-AAU11054 and AAU11521-AAU11537, used in the invention to improve
CC  the ring-expanding activity of expandase. Note: The present protein
CC  sequence is not shown in the specification but is derived from the
CC  Streptomyces clavuligerus wild-type penicillin N expandase sequence given
CC  in Figure 1 (see AAU11044)
XX

```

SQ Sequence 311 AA;

Query Match 99.7%; Score 1631; DB 5; Length 311;
Best Local Similarity 99.4%; Pred. No. 2.2e-167;
Matches 309; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDTVPFSLAELOQGLHODEFRRLDKGLFYLTDCGLTDTLKSADIVIDPFHSGE 60
DB 1 MDTVPFSLAELOQGLHODEFRRLDKGLFYLTDCGLTDTLKSADIVIDPFHSGE 60
QY 61 AEKRAVTSVPVPMRGFTGLSESTAOITNTGSDYSCYMGTAADNLFPSSGDFER1WT 120
DB 61 AEKRAVTSVPVPMRGFTGLSESTAOITNTGSDYSCYMGTAADNLFPSSGDFER1WT 120
QY 121 QYFDROYTASRAVAAREVLATGTEPDGVEAFLDCEPLLRFRYPQVPEHRSABEQPIRM 180
DB 121 QYFDROYTASRAVAAREVLATGTEPDGVEAFLDCEPLLRFRYPQVPEHRSABEQPIRM 180
QY 181 APHYDLMSVTLIQOTPCANGFVSLQAEVGAFTDLPRPDVAVFCGAIATLVGGQVKA 240
DB 181 APHYDLMSVTLIQOTPCANGFVSLQAEVGAFTDLPRPDVAVFCGAIATLVGGQVKA 240
QY 241 PRHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
DB 241 PRHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
QY 301 NYVINIRRTSKA 311
DB 301 NYVINIRRTSKA 311

RESULT 7
AAU11053
ID AAU11053 standard; protein; 310 AA.
XX AAU11053;
XX
DT 12-MAR-2002 (first entry)
XX
DE S. clavuligerus penicillin N expandase 1 residue deletion at C-terminus.
XX
KW Penicillin N expandase; ring-expanding; penicillin G;
KW phenylacetyl-7-ADCA; amino desacetoxycephalosporanic acid; cephalosporin;
KW cephalaxin; penicillin V; mutant; mutein.
XX
OS Streptomycetes clavuligerus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 310 /note= "Protein is truncated at this residue"
XX
XX WO200185951-A1.
XX
XX 15-NOV-2001.
XX
XX 09-MAY-2001; 2001MO-GB002047.
XX
XX 09-MAY-2000; 2000GB-00011185.
XX
XX (ACSD-) ACS DOBFAR UK LTD.
XX
XX Johnson RI, Newbert RW;
XX
XX WPI; 2002-075247/10.
XX
XX
XX New enzyme useful for ring-expanding penicillin G to produce phenylacetyl
XX PT -7-amino desacetoxycephalosporanic acid, comprising penicillin expandase
XX PT having increased specificity for substrates such as penicillin G.
XX
XX Disclosure; Page; 42pp; English.
XX
XX The present invention relates to new penicillin expandases modified to

CC Improve the ring-expanding activity for a substrate which is not the
CC natural substrate of the unmodified expandase. The invention is useful
CC for ring-expansion of penicillin G to produce phenylacetyl-7-ADCA (amino
CC desacetoxycephalosporanic acid) and the phenylacetyl side chain is
CC removed from phenylacetyl-7-ADCA to produce 7-ADCA, which is useful as a
CC starting point for the production of a range of semi-synthetic
CC cephalosporins, most notably cephalaxin. The polynucleotide of the
CC invention is useful for producing a primer e.g. a PCR primer and as a
CC probe. The modification of the penicillin expandase enhances the activity
CC of the enzyme such as penicillin N expandase for penicillin G or V as a
CC substrate. The modified expandase has enhanced catalytic activity or
CC increased specificity for another substrate such as penicillin G. This
CC sequence is one of a collection of Streptomycetes clavuligerus mutants,
CC AAU1046-AAU1054 and AAU1521-AAU1537, used in the invention to improve
CC the ring-expanding activity of expandase. Note: The present protein
CC sequence is not shown in the specification but is derived from the
CC Streptomycetes clavuligerus wild-type penicillin N expandase sequence given
CC in figure 1 (see AAU11044)

SQ Sequence 310 AA;

Query Match 99.6%; Score 1630; DB 5; Length 310;
Best Local Similarity 99.7%; Pred. No. 2.8e-167;
Matches 309; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDTVPFSLAELOQGLHODEFRRLDKGLFYLTDCGLTDTLKSADIVIDPFHSGE 60
DB 1 MDTVPFSLAELOQGLHODEFRRLDKGLFYLTDCGLTDTLKSADIVIDPFHSGE 60
QY 61 AEKRAVTSVPVPMRGFTGLSESTAOITNTGSDYSCYMGTAADNLFPSSGDFER1WT 120
DB 61 AEKRAVTSVPVPMRGFTGLSESTAOITNTGSDYSCYMGTAADNLFPSSGDFER1WT 120
QY 121 QYFDROYTASRAVAAREVLATGTEPDGVEAFLDCEPLLRFRYPQVPEHRSABEQPIRM 180
DB 121 QYFDROYTASRAVAAREVLATGTEPDGVEAFLDCEPLLRFRYPQVPEHRSABEQPIRM 180
QY 181 APHYDLMSVTLIQOTPCANGFVSLQAEVGAFTDLPRPDVAVFCGAIATLVGGQVKA 240
DB 181 APHYDLMSVTLIQOTPCANGFVSLQAEVGAFTDLPRPDVAVFCGAIATLVGGQVKA 240
QY 241 PRHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
DB 241 PRHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
QY 301 NYVINIRRTSK 310
DB 301 NYVINIRRTSK 310

RESULT 8
AAU11525
ID AAU11525 standard; protein; 310 AA.
XX AAU11525;
XX
DT 12-MAR-2002 (first entry)
XX
DE S. clavuligerus penicillin N expandase K310 deletion mutant.
XX
KW Penicillin N expandase; ring-expanding; penicillin G;
KW phenylacetyl-7-ADCA; amino desacetoxycephalosporanic acid; cephalosporin;
KW cephalaxin; penicillin V; mutant; mutein.
XX
OS Streptomycetes clavuligerus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 310 /note= "Protein is truncated at this residue"
XX
XX WO200185951-A1.
XX

PD 15-NOV-2001.
XX 09-MAY-2001; 2001WO-GB002047.
XX 09-MAY-2000; 2000GB-0001185.
PR (ACSD-) ACS DOBFAR UK LTD.
XX
PI Johnson RI, Newbert RW;
XX WPI; 2002-075247/10.
XX
PT New enzyme useful for ring-expanding penicillin G to produce phenylacetyl
PT -7-amino deacetoxycephalosporanic acid, comprises penicillin expandase
PT having increased specificity for substrates such as penicillin G.
XX
PS Example; Page; 42pp; English.
XX
XX The present invention relates to new penicillin expandases modified to
CC improve the ring-expanding activity for a substrate which is not the
CC natural substrate of the unmodified expandase. The invention is useful
CC for ring-expansion of penicillin G to produce phenylacetyl-7-ADCA (amino
CC deacetoxycephalosporanic acid) and the phenylacetyl side chain is
CC removed from phenylacetyl-7-ADCA to produce 7-ADCA, which is useful as a
CC starting point for the production of a range of semi-synthetic
CC cephalosporins, most notably cephalixin. The polynucleotide of the
CC invention is useful for producing a primer e.g. a PCR primer and as a
CC probe. The modification of the penicillin expandase enhances the activity
CC of the enzyme such as penicillin N expandase for penicillin G or V as a
CC substrate. The modified expandase has enhanced catalytic activity or
CC increased specificity for another substrate such as penicillin G. This
CC sequence is one of a collection of Streptomyces clavuligerus mutants,
CC AAU1046-AAU1054 and AAU1521-AAU1537, used in the invention to improve
CC the ring-expanding activity of expandase. Note: The present protein
CC sequence is not shown in the specification but is derived from the
CC Streptomyces clavuligerus wild-type penicillin N expandase sequence given
CC in figure 1 (see AAU1044)
XX
SQ Sequence 310 AA;

Query Match 99.6%; Score 1630; DB 5; Length 310;
Best Local Similarity 99.7%; Pred. No. 2.8e-167;
Matches 309; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDTTVPFSLAEIQGLHODEFRRLRDGKLFYLTDCGLTDTLKSADIVIDFFHSGS 60
DB 1 MDTTVPFSLAEIQGLHODEFRRLRDGKLFYLTDCGLTDTLKSADIVIDFFHSGS 60
QY 61 AEKRAVTSVPVTRMRGFTGLSESTAOITNTGSDYSYSCMSGTAADNLPSSGDFERIMT 120
DB 61 AEKRAVTSVPVTRMRGFTGLSESTAOITNTGSDYSYSCMSGTAADNLPSSGDFERIMT 120
QY 121 QYPRQYTASRAVAREVLATGTEPDGVEAFIDCEPLLRFRYPQVPEHRSABEQPLRM 180
DB 121 QYPRQYTASRAVAREVLATGTEPDGVEAFIDCEPLLRFRYPQVPEHRSABEQPLRM 180
QY 181 APHYDLSMWTLLIQOTPCANGFVSLQAEVGAFTDLPYRDAVYVFCGAIATLVTGGQVKA 240
DB 181 APHYDLSMWTLLIQOTPCANGFVSLQAEVGAFTDLPYRDAVYVFCGAIATLVTGGQVKA 240
QY 241 PRHHVAAPRRDOJAGSSRTSSVFELRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
DB 241 PRHHVAAPRRDOJAGSSRTSSVFELRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
QY 301 NYVNIIRRTSK 310
DB 301 NYVNIIRRTSK 310

RESULT 9
AAP93215
ID AAP93215 standard; protein; 311 AA.
XX

AC AAP93215;
XX 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 30-MAR-1990 (first entry)
XX
DE Deacetoxycephalosporin C synthetase.
XX
KM Deacetoxycephalosporin C synthetase; DAOCs; cephalosporin; penicillin N;
KM antibiotics.
XX
OS Streptomyces clavuligerus.
XX
PN BP341892-A.
XX
PD 15-NOV-1989.
XX
XX 09-MAY-1988; 88US-00192273.
XX
PR 09-MAY-1988; 88US-00192273.
XX
PA (ELIL) LILLY & CO ELI.
XX
PI Ingolia TD, Kovacevic S, Miller JR, Skatrud PL;
XX WPI; 1989-334231/46.
DR N-PSDB; AAN92262.
XX
XX New recombinant DNA encoding de-acetoxy-cephalosporin C synthetase - for
PT increasing or inducing cephalosporin synthesis in microorganisms or in
PT vitro.
XX
PS Claim 2; Page 26-27; 46pp; English.
XX
XX The sequence encodes deacetoxycephalosporin C synthetase (DAOCs) which
CC catalyses expansion of penicillin N. Cephalosporium strains transformed
CC with vectors carrying the gene will produce antibiotics more efficiently.
CC (Updated on 25-MAR-2003 to correct PA field.) (Updated on 27-AUG-2003 to
CC correct OS field.)
XX
SQ Sequence 311 AA;

Query Match 99.6%; Score 1630; DB 1; Length 311;
Best Local Similarity 99.7%; Pred. No. 2.9e-167;
Matches 310; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDTTVPFSLAEIQGLHODEFRRLRDGKLFYLTDCGLTDTLKSADIVIDFFHSGS 60
DB 1 MDTTVPFSLAEIQGLHODEFRRLRDGKLFYLTDCGLTDTLKSADIVIDFFHSGS 60
QY 61 AEKRAVTSVPVTRMRGFTGLSESTAOITNTGSDYSYSCMSGTAADNLPSSGDFERIMT 120
DB 61 AEKRAVTSVPVTRMRGFTGLSESTAOITNTGSDYSYSCMSGTAADNLPSSGDFERIMT 120
QY 121 QYPRQYTASRAVAREVLATGTEPDGVEAFIDCEPLLRFRYPQVPEHRSABEQPLRM 180
DB 121 QYPRQYTASRAVAREVLATGTEPDGVEAFIDCEPLLRFRYPQVPEHRSABEQPLRM 180
QY 181 APHYDLSMWTLLIQOTPCANGFVSLQAEVGAFTDLPYRDAVYVFCGAIATLVTGGQVKA 240
DB 181 APHYDLSMWTLLIQOTPCANGFVSLQAEVGAFTDLPYRDAVYVFCGAIATLVTGGQVKA 240
QY 241 PRHHVAAPRRDOJAGSSRTSSVFELRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
DB 241 PRHHVAAPRRDOJAGSSRTSSVFELRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
QY 301 NYVNIIRRTSKA 311
DB 301 NYVNIIRRTSKA 311

RESULT 10
AAM94168

```

ID  AAW34168 standard; protein; 311 AA.
XX
XX  AAW34168;
AC
XX  17-FEB-1998 (first entry)
DT
XX  Expandase mutant R74Q.
DE
XX  Expandase; 7-aminodesacetoxycephalosporanic acid; 7-ADCA production;
KM  thiazolidine ring expansion; penicillin N; dihydrothiazine ring; enzyme;
KM  semi-synthetic cephalosporin production; penicillin G; isopenicillin N;
XX  cephalosporin ring; DOAC; mutant.
XX
OS  Streptomyces clavuligerus.
OS  Synthetic.
XX
XX  Key      Location/Qualifiers
FH  Misc-difference 74
FT  /note= "Arg to Gln mutation"
XX
XX  MO9720053-A2.
XX
XX  05-JUN-1997.
PD
XX
XX  27-NOV-1996; 96WO-EP005358.
XX
XX  27-NOV-1995; 95EP-00203259.
XX
XX  27-NOV-1995; 95US-0007564P.
XX
XX  (KONN ) GIST-BROCADES BV.
XX  (UYOX-) UNIV OXFORD.
XX
XX  Sutherland JD, Bovenberg RAL, Van Der Laan JM;
XX  WPI; 1997-310608/28.
XX
XX  Expandase enzymes with mutation(s) to alter substrate specificity - used
XX  particularly for production of 7-amino:desacetoxy:cephalosporanic acid.
XX
XX  Claim 2; Page; 21pp; English.
XX
XX  AAW34166-W34177 represent mutant expandase sequences of the invention.
XX  These sequences represent mutated versions of the Streptomyces
XX  clavuligerus expandase sequences shown in AAW17775. The DNA encoding the
XX  wild type expandase was subjected to site directed mutagenesis to create
XX  these sequences. The mutant expandase sequences have an altered substrate
XX  specificity compared to the wild type expandase. Expandase catalyses the
XX  expansion of the 5-membered thiazolidine ring of penicillin N to the 6-
XX  membered dihydrothiazine ring of DOAC. The mutant expandase gene is used
XX  particularly for the production of 7-aminodesacetoxycephalosporanic acid
XX  (7-ADCA). 7-ADCA is one of the key intermediates that is used in the
XX  preparation of semi-synthetic cephalosporins. These sequences have
XX  improved ability to expand penicillin G into a cephalosporin ring in
XX  vitro and/or in vivo where other penicillins such as penicillin N and
XX  isopenicillin N can act as competing substrates
XX
XX  Sequence 311 AA;
SQ

Query Match      99.6%; Score 1630; DB 2; Length 311;
Best Local Similarity 99.4%; Pred. No. 2,9e-167;
Matches 309; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  181 APHYDLSWVTLIQOTPCANGFVSLQAEVGAFTDLPRPADAVIVFCGAIATLVYGGGYKA 240
DB  181 APHYDLSWVTLIQOTPCANGFVSLQAEVGAFTDLPRPADAVIVFCGAIATLVYGGGYKA 240
QY  241 PRHHVAAPRRDQIAGSSRSTSSVFFFLRPVADPTFVSPLARECGFVSLDGETATPDWIGG 300
DB  241 PRHHVAAPRRDQIAGSSRSTSSVFFFLRPVADPTFVSPLARECGFVSLDGETATPDWIGG 300
QY  301 NYVNIIRRTSKA 311
DB  301 NYVNIIRRTSKA 311

RESULT 11
AAW34176
ID  AAW34176 standard; protein; 311 AA.
XX
XX  AAW34176;
AC
XX  17-FEB-1998 (first entry)
DT
XX  Expandase mutant R266Q.
DE
XX
XX  Expandase; 7-aminodesacetoxycephalosporanic acid; 7-ADCA production;
KM  thiazolidine ring expansion; penicillin N; dihydrothiazine ring; enzyme;
KM  semi-synthetic cephalosporin production; penicillin G; isopenicillin N;
XX  cephalosporin ring; DOAC; mutant.
XX
OS  Streptomyces clavuligerus.
OS  Synthetic.
XX
XX  Key      Location/Qualifiers
FH  Misc-difference 266
FT  /note= "Arg to Gln mutation"
XX
XX  MO9720053-A2.
XX
XX  05-JUN-1997.
PD
XX
XX  27-NOV-1996; 96WO-EP005358.
XX
XX  27-NOV-1995; 95EP-00203259.
XX
XX  27-NOV-1995; 95US-0007564P.
XX
XX  (KONN ) GIST-BROCADES BV.
XX  (UYOX-) UNIV OXFORD.
XX
XX  Sutherland JD, Bovenberg RAL, Van Der Laan JM;
XX  WPI; 1997-310608/28.
XX
XX  Expandase enzymes with mutation(s) to alter substrate specificity - used
XX  particularly for production of 7-amino:desacetoxy:cephalosporanic acid.
XX
XX  Claim 2; Page; 21pp; English.
XX
XX  AAW34166-W34177 represent mutant expandase sequences of the invention.
XX  These sequences represent mutated versions of the Streptomyces
XX  clavuligerus expandase sequences shown in AAW17775. The DNA encoding the
XX  wild type expandase was subjected to site directed mutagenesis to create
XX  these sequences. The mutant expandase sequences have an altered substrate
XX  specificity compared to the wild type expandase. Expandase catalyses the
XX  expansion of the 5-membered thiazolidine ring of penicillin N to the 6-
XX  membered dihydrothiazine ring of DOAC. The mutant expandase gene is used
XX  particularly for the production of 7-aminodesacetoxycephalosporanic acid
XX  (7-ADCA). 7-ADCA is one of the key intermediates that is used in the
XX  preparation of semi-synthetic cephalosporins. These sequences have
XX  improved ability to expand penicillin G into a cephalosporin ring in
XX  vitro and/or in vivo where other penicillins such as penicillin N and
XX  isopenicillin N can act as competing substrates
XX
XX  Sequence 311 AA;
SQ

```

Query Match 99.6%; Score 1630; DB 2; Length 311;
Best Local Similarity 99.4%; Pred. No. 2.9e-167;
Matches 309; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MDTTPTFSLAELOQGLHODEFRRCIRDKGLFYLTDCGLTDTLKSADIVIDFEHSGE 60
Db 1 MDTTPTFSLAELOQGLHODEFRRCIRDKGLFYLTDCGLTDTLKSADIVIDFEHSGE 60
QY 61 AEKRAVTSVPPTMRKRGFTGLESESTAOITNTGYSYSDYSCYSGTADNLFPSGDFERIWT 120
Db 61 AEKRAVTSVPPTMRKRGFTGLESESTAOITNTGYSYSDYSCYSGTADNLFPSGDFERIWT 120
QY 121 QYFDRQYTSRAVARAVLRATGTEPDGVEAFIDCEPLRFRFPQVPEHRSABEQPLRM 180
Db 121 QYFDRQYTSRAVARAVLRATGTEPDGVEAFIDCEPLRFRFPQVPEHRSABEQPLRM 180
QY 181 APHYDLSMTWTLIQOTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVTGQVKA 240
Db 181 APHYDLSMTWTLIQOTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVTGQVKA 240
QY 241 PRHHVAAPRRDQIAGSSRTSSVFELRPNDFTFSVPLARECGDFVSLDGETATFQDWIGG 300
Db 241 PRHHVAAPRRDQIAGSSRTSSVFELRPNDFTFSVPLARECGDFVSLDGETATFQDWIGG 300
QY 301 NYVINIRRTSKA 311
Db 301 NYVINIRRTSKA 311
```

RESULT 12
AAW40424
ID AAW40424 standard; protein; 311 AA.

AC AAW40424;

DT 02-JUL-1998 (first entry)

DE S. clavuligerus expandase protein.

KW Expandase; adipoyl cephalosporin; desacetoxycephalosporin C synthase;

KM DAOCS; 5-carboxypentanoyl cephalosporin.

OS Streptomyces clavuligerus.

PN MO9802551-A2.

PD 22-JAN-1998.

PF 15-JUL-1997; 97WO-EP003879.

PR 16-JUL-1996; 96EP-00201988.

PA (KONN) GIST-BROCADES BV.

PI Bovenberg RAL, Van Der laan JM, Kerkman R, Nieboer M;

DR WPI; 1998-110597/10.

PT Modified expandase genes - which encode mutant expandase with amino acid
substitutions used for the production of cephalosporin compounds.

PS Disclosure; Page; 48pp; English.

XX This sequence represents the expandase protein (also known as
XX desacetoxycephalosporin C synthase or DAOCS) from Streptococcus
XX clavuligerus. Mutants of this protein are used in a novel method
XX involving a biosynthetic process for preparation and recovery of adipoyl
XX cephalosporins (5-carboxypentanoyl cephalosporins). The modified
XX expandase provides for the more efficient production of cephalosporins
XX Sequence 311 AA;

Query Match 99.6%; Score 1630; DB 2; Length 311;
Best Local Similarity 99.7%; Pred. No. 2.9e-167;
Matches 310; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 MDTTPTFSLAELOQGLHODEFRRCIRDKGLFYLTDCGLTDTLKSADIVIDFEHSGE 60
Db 1 MDTTPTFSLAELOQGLHODEFRRCIRDKGLFYLTDCGLTDTLKSADIVIDFEHSGE 60
QY 61 AEKRAVTSVPPTMRKRGFTGLESESTAOITNTGYSYSDYSCYSGTADNLFPSGDFERIWT 120
Db 61 AEKRAVTSVPPTMRKRGFTGLESESTAOITNTGYSYSDYSCYSGTADNLFPSGDFERIWT 120
QY 121 QYFDRQYTSRAVARAVLRATGTEPDGVEAFIDCEPLRFRFPQVPEHRSABEQPLRM 180
Db 121 QYFDRQYTSRAVARAVLRATGTEPDGVEAFIDCEPLRFRFPQVPEHRSABEQPLRM 180
QY 181 APHYDLSMTWTLIQOTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVTGQVKA 240
Db 181 APHYDLSMTWTLIQOTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVTGQVKA 240
QY 241 PRHHVAAPRRDQIAGSSRTSSVFELRPNDFTFSVPLARECGDFVSLDGETATFQDWIGG 300
Db 241 PRHHVAAPRRDQIAGSSRTSSVFELRPNDFTFSVPLARECGDFVSLDGETATFQDWIGG 300
QY 301 NYVINIRRTSKA 311
Db 301 NYVINIRRTSKA 311
```

RESULT 13
AAW34175
ID AAW34175 standard; protein; 311 AA.

AC AAW34175;

DT 17-FEB-1998 (first entry)

DE Expandase mutant R266N.

KW Expandase; 7-aminodesacetoxycephalosporanic acid; 7-ADCA production;

KM thiazolidine ring expansion; penicillin N; dihydrothiazine ring; enzyme;

KM semi-synthetic cephalosporin production; penicillin G; isopenicillin N;

KM cephalosporin ring; DOAC; mutant.

OS Streptomyces clavuligerus.

PN Synthetic.

PF MO9720053-A2.

PD 05-JUN-1997.

PF 27-NOV-1996; 96WO-EP005358.

PR 27-NOV-1995; 95EP-00203259.

PR 27-NOV-1995; 95US-0007564P.

PA (KONN) GIST-BROCADES BV.

PA (UYOX-) UNIV OXFORD.

PI Sutherland JD, Bovenberg RAL, Van Der laan JM;

DR WPI; 1997-310608/28.

PT Expandase enzymes with mutation(s) to alter substrate specificity - used
PT particularly for production of 7-amino-desacetoxycephalosporanic acid.

XX Claim 2; Page; 21pp; English.

XX AAW34166-W34177 represent mutant expandase sequences of the invention.

CC These sequences represent mutated versions of the Streptomyces
CC clavuligenus expandase sequences shown in AAM17775. The DNA encoding the
CC wild type expandase was subjected to site directed mutagenesis to create
CC these sequences. The mutant expandase sequences have an altered substrate
CC specificity compared to the wild type expandase. Expandase catalyses the
CC expansion of the 5-membered thiazolidine ring of penicillin N to the 6-
CC membered dihydrothiazine ring of DOAC. The mutant expandase gene is used
CC particularly for the production of 7-aminodeacetoxycephalosporanic acid
CC (7-ADCA). 7-ADCA is one of the key intermediates that is used in the
CC preparation of semi-synthetic cephalosporins. These sequences have
CC improved ability to expand penicillin G into a cephalosporin ring in
CC vitro and/or in vivo where other penicillins such as penicillin N and
CC isopenicillin N can act as competing substrates

SQ Sequence 311 AA;

Query Match	99.6%	Score 1629	DB 2	Length 311
Best Local Similarity	99.4%	Pred. No. 3.7e-167		
Matches 309	Conservative 1	Mismatches 1	Indels 0	Gaps 0

Oy	1	MDTVPPTSLAELOOG:HODEFERRCLRDKGLYLTDCGLTDLIELSAKXIVIDFFPHSGE	60
		MDTVPPTSLAELOOG:HODEFERRCLRDKGLYLTDCGLTDLIELSAKXIVIDFFPHSGE	60
Db	1	MDTVPPTSLAELOOG:HODEFERRCLRDKGLYLTDCGLTDLIELSAKXIVIDFFPHSGE	60
Oy	61	AEKRAVTSVPVPMRRGFTGLSEESTAOITNTGTSYSDYSKCYSMGTADNLFPSGDFERRINT	120
		AEKRAVTSVPVPMRRGFTGLSEESTAOITNTGTSYSDYSKCYSMGTADNLFPSGDFERRINT	120
Db	61	AEKRAVTSVPVPMRRGFTGLSEESTAOITNTGTSYSDYSKCYSMGTADNLFPSGDFERRINT	120

Qy 12 QYFBEQYTASRAVAREVLATGTETPEPGGVEAFIDCEPLLRRYFPQVPEHNSAEEOPLRM 180

Ddb 121 QYFDRQYTASRAVAREVLKRTGTETPDGGVEAFIDCEPLLRKYFPQVPEHNSAEEOPLRM 180

QY 181 APHHDLSMYLLIQQTPCANGFVSLQAEVGGAFTDLYRPRDAVLVEGAIATILVTGGQVKA 240

Db 181 APHHDLSMYLLIQQTPCANGFVSLQAEVGGAFTDLYRPRDAVLVEGAIATILVTGGQVKA 240

0Y PRHHVAAPRRDQIAGSSRTSSVFELRNPAEFTFSVPLAREGEGDVSLDGEIATFDQWIGG 300
 241 PRHHVAAPRRDQIAGSSRTSSVFELRNPAEFTFSVPLAREGEGDVSLDGEIATFDQWIGG 300

QY	301 NYVNIKRTSKA	311
Db	301 NYVNIKRTSKA	311

RESULT 14
AAE12016
ID AAE12016 standard: protein: 311 AA

AA	AAE12016;
AC	
XX	
DT	18-DEC-2001 (first entry)

DE Streptomyces clavuligerus modified deacetoxycephalosporin C synthetase.

KW Deacetoxycephalosporin C synthetase; DAOCS; antibiotic; penicillin;
KW oxygenase: non-haeme iron (II) dependent family: oxidase: mutant: mutant

Synthetic.

	Key	Location/Qualifiers
FH	Misc-difference	181
FT		/note= "Wild type A1a substituted with Arg"
FT		

PN US6284483-B1.

PD 04-SEP-200

PF 06-OCT-1999; 99US-00413231.

PR 06-OCT-1999; 99US-00413231.
XX

PA (UNMS) UNIV MICHIGAN STATE.

PI Dilley DR, Kadyrzhanova DK, Wang Z, Warner TM, VY

DR WPI; 2001-615433/71

PT Method for producing antibiotics, particularly penicillin G or V, PT comprises employing a modified isopenicillin N synthetase, in either an organism or a cell-free system under the control of bicarbonate.

PS Disclosure; Col 45-48; 27pp; English.

CC The invention relates to a modified enzyme of a non-haem iron (II)
CC dependent family of oxygenases and oxidases which renders the enzyme
CC dependent on bicarbonate for activity. The invention also related to a
CC method for producing penicillin G or V comprises employing a modified
CC enzyme, particularly isopenicillin N synthetase (IPNS), in either an
CC organism or a cell-free system, under the control of bicarbonate. The
CC modified enzymes are IPNS, deacetoxycephalosporin C synthetase (DAOCS)
CC and deacetylcephalosporin C synthetase (DACSC). The method is useful for
CC producing antibiotics, particularly penicillin G or V. The method is
CC useful for making organisms useful for making an antibiotic dependent
CC bicarbonate to make the antibiotic. The present sequence is Streptomyces
CC clavuligerus modified deacetoxycephalosporin C synthetase showing an
CC Arg181 in place of Ala181

SQ Sequence 311 AA;

Query Match	99.6%	Score 1629	DB 4	Length 311
Best Local Similarity	99.4%	Pred. No. 3.7e-167		
Matches 309	Conservative 1	Mismatches 1	Indels 0	Gaps 0

```
QY      1 MDTVPTFTSLAELQQGSHQDERRRCRLDKGLFLLTDCGILTDELKSAKDIVIDFEHGSE 60
        |||||
DB       1 MDTVPFTSLAELQQGSHQDERRRCRLDKGLFYLTDCGILTDTEKSAKDVIDFEHGE 60
```

Qy 61 AEKRAVTSPVPTMRGFTGLSESTADITNTGSSYSDYMCYMGATADNLPSSGDERIWT 12
 |||||
 Db 61 AEKRAVTSPVPTMRGFTGLSESTADITNTGSSYSDYMCYMGATADNLPSSGDERIWT 12

```

Qy      121 QYFDRQYTA$AVAREVLRATGTEPDG$VEAFLDCEPLRFRYPQVPEHRS$AEQPLRM 18
         |||
Db      121 QYFDRQYTA$AVAREVLRATGTEPDG$VEAFLDCEPLRFRYPQVPEHRS$AEQPLRM 18

```

QY 181 APHYDLSMTLIIQQTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCGATLTLVIGGQVKA 24
 |||||
 Db 181 RPHYDLSMTLIIQQTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCGATLTLVIGGQVKA 24
 |||||

QY 241 PRHNVAAPRKQDIAGSSRTSSVFLEKRNADFTTSVPLARECGFDVSLDGETATFDQWIGG 30

Db 241 PRHNVAAPRKQDIAGSSRTSSVFLEKRNADFTTSVPLARECGFDVSLDGETATFDQWIGG 30

Qy	301	NYVNI	IRTSKA	311
Db	301	NYVNI	IRTSKA	311

RESULT 15
AAU11047
ID AAU11047 standard; protein; 311 AA

AC AAU11047;

XX Streptomyces clavuligerus penicillin N expandase L158A mutant.

KM phenylacetyl-7-ADCA; amino desacetoxyccephalosporanic acid; cephalosporin;

XX

9

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2005, 15:02:23 ; Search time 31 Seconds
(without alignments)
965.271 Million cell updates/sec

Title: 10719236-1_232-1164

Perfect score: 1636

Sequence: 1 MDTVPPTFSLAELOQGLHQP.....ATFDWMIGNVNIRRTSKA 311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1636	100.0	311	2 A32043	deacetoxycephalosporin C synthetase - Streptomyces clavuligerus
2	1634	99.9	311	2 TS2312	deacetoxycephalosporin C synthetase - Streptomyces clavuligerus
3	1195.5	73.1	314	2 S30900	deacetoxycephalosporin C synthetase - Streptomyces clavuligerus
4	972.5	59.4	310	2 S40253	deacetoxycephalosporin C synthetase - Streptomyces clavuligerus
5	946.5	57.9	318	2 A39204	deacetoxycephalosporin C synthetase - Streptomyces clavuligerus
6	944	57.7	319	2 S54100	deacetoxycephalosporin C synthetase - Streptomyces clavuligerus
7	933.5	57.1	332	2 A29711	deacetoxycephalosporin C synthetase - Streptomyces clavuligerus
8	859	52.5	313	2 S54101	deacetoxycephalosporin C synthetase - Streptomyces clavuligerus
9	178	10.9	329	2 A61155	isopenicillin N synthetase - Streptomyces clavuligerus
10	168	10.3	329	2 A28894	isopenicillin N synthetase - Streptomyces clavuligerus
11	164	10.0	326	2 S08218	isopenicillin N synthetase - Streptomyces clavuligerus
12	164	10.0	326	2 S54099	isopenicillin N synthetase - Streptomyces clavuligerus
13	147	9.0	333	2 B32344	isopenicillin N synthetase - Streptomyces clavuligerus
14	144.5	8.8	358	2 C84713	probable dihydroxyflavonol synthase - Streptomyces clavuligerus
15	143	8.7	348	2 S3510	probable dihydroxyflavonol synthase - Streptomyces clavuligerus
16	142	8.7	328	2 S15284	isopenicillin N synthetase - Streptomyces clavuligerus
17	139.5	8.5	362	2 D84713	probable dihydroxyflavonol synthase - Streptomyces clavuligerus
18	137.5	8.4	349	2 T07373	probable gibberellin 20-oxi leucanthocyanidin synthetase - Streptomyces clavuligerus
19	134.5	8.2	376	2 D96635	probable gibberellin 20-oxi leucanthocyanidin synthetase - Streptomyces clavuligerus
20	133.5	8.2	331	2 A28467	isopenicillin N synthetase - Streptomyces clavuligerus
21	129.5	7.9	331	2 S04441	isopenicillin N synthetase - Streptomyces clavuligerus
22	129	7.9	321	2 A58458	isopenicillin N synthetase - Streptomyces clavuligerus
23	128.5	7.9	355	2 A50160	probable iron/asco gibberellin 20-oxi leucanthocyanidin synthetase - Streptomyces clavuligerus
24	124	7.7	380	2 T48491	isopenicillin N synthetase - Streptomyces clavuligerus
25	124	7.6	371	2 T08008	isopenicillin N synthetase - Streptomyces clavuligerus
26	123	7.5	339	2 T05903	isopenicillin N synthetase - Streptomyces clavuligerus
27	122.5	7.5	331	2 A27355	isopenicillin N synthetase - Streptomyces clavuligerus
28	122.5	7.5	386	2 T09675	probable gibberellin 20-oxi leucanthocyanidin synthetase - Streptomyces clavuligerus
29	122.5	7.5	386	2 T09664	probable gibberellin 20-oxi leucanthocyanidin synthetase - Streptomyces clavuligerus

30	122	7.5	356	2 T05119	leucanthocyanidin synthetase - Streptomyces clavuligerus
31	119.5	7.3	358	2 S44251	gibberellin 20-oxi leucanthocyanidin synthetase - Streptomyces clavuligerus
32	119.5	7.2	370	2 T11849	gibberellin 20-oxi leucanthocyanidin synthetase - Streptomyces clavuligerus
33	117.5	7.2	338	2 S47972	gibberellin 20-oxi leucanthocyanidin synthetase - Streptomyces clavuligerus
34	117.5	7.2	356	2 T05551	gibberellin 20-oxi leucanthocyanidin synthetase - Streptomyces clavuligerus
35	117	7.2	382	2 T11847	gibberellin 20-oxi leucanthocyanidin synthetase - Streptomyces clavuligerus
36	116	7.1	320	2 C83628	probable oxidoreductase - Streptomyces clavuligerus
37	115	7.0	365	2 S31921	probable oxidoreductase - Streptomyces clavuligerus
38	114.5	7.0	321	2 T41002	hypothetical protein - Streptomyces clavuligerus
39	114	7.0	298	2 T09145	hypothetical protein - Streptomyces clavuligerus
40	114	7.0	352	2 T01606	probable flavonol synthase - Streptomyces clavuligerus
41	114	7.0	356	2 T07972	leucanthocyanidin synthetase - Streptomyces clavuligerus
42	112.5	6.9	370	2 T47932	1-aminocyclopropan carboxylate oxidase - Streptomyces clavuligerus
43	111	6.8	363	2 T49209	leucanthocyanidin synthetase - Streptomyces clavuligerus
44	110.5	6.8	306	2 T45586	hypothetical protein - Streptomyces clavuligerus
45	110.5	6.8	377	2 T10222	gibberellin 20-oxi leucanthocyanidin synthetase - Streptomyces clavuligerus

ALIGNMENTS

```
RESULT 1
A32043
deacetoxycephalosporin C synthetase - Streptomyces clavuligerus
C:Species: Streptomyces clavuligerus
C:Date: 13-Jul-1989 #sequence_revision 13-Jul-1989 #text_change 19-May-2000
C:Accession: A32043
R:Kovacevic, S.; Weigel, B.J.; Tobin, M.B.; Ingolia, T.D.; Miller, J.R.
J. Bacteriol. 171, 754-760, 1989
A:Title: Cloning, characterization, and expression in Escherichia coli of the Streptomyces
A:Reference number: A32043; MIMD:89123150; PMID:2644235
A:Accession: A32043
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-311 <KOV>
C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase

Query Match      100.0%; Score 1636; DB 2; Length 311;
Best Local Similarity 100.0%; Pred. No. 3.1e-139;
Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MDTVPPTFSLAELOQGLHQP...ATFDWMIGNVNIRRTSKA 60
DB 1 MDTVPPTFSLAELOQGLHQP...ATFDWMIGNVNIRRTSKA 60
OY 61 AEKRAVTSVPVPMRGRFTGLESESTAOITNTGYSYDYSWCYSMGTADNLPSPGDFERIWT 120
DB 61 AEKRAVTSVPVPMRGRFTGLESESTAOITNTGYSYDYSWCYSMGTADNLPSPGDFERIWT 120
OY 121 QYPRQYTSARAVRREVLRAATGTEPPDGVAFPLDCEPLRPFRVPOVPEHRSASEQPLRM 180
DB 121 QYPRQYTSARAVRREVLRAATGTEPPDGVAFPLDCEPLRPFRVPOVPEHRSASEQPLRM 180
OY 181 APHFDLSMTVLIQOTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCAGATLVYGGGVKA 240
DB 181 APHFDLSMTVLIQOTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCAGATLVYGGGVKA 240
OY 241 PRHHVAAPRRDQJAGSSRTSSVFPLRPNADFTFESVPLARECGFVSLDGETATFQDWIGG 300
DB 241 PRHHVAAPRRDQJAGSSRTSSVFPLRPNADFTFESVPLARECGFVSLDGETATFQDWIGG 300
OY 301 NYVNIIRRTSKA 311
DB 301 NYVNIIRRTSKA 311

RESULT 2
TS2312
deacetoxycephalosporin C synthetase [imported] - Streptomyces clavuligerus
C:Species: Streptomyces clavuligerus
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 09-Jul-2004
C:Accession: TS2312
R:Kovacevic, S.; Tobin, M.B.; Miller, J.R.
```

J. Bacteriol. 172, 3952-3958, 1990
A>Title: The beta-lactam biosynthesis genes for isopenicillin N epimerase and deacetoxy
A;Reference number: Z26033; MUID:90299822; PMID:1694525
A;Accession: T52312
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-311 <KOV>
A;Cross-references: UNIPROT:P18548; EMBL:M32324; PIDD:AAA6715.1
C;Genetics:
A;Gene: cefE
C;Superfamily: isopenicillin N synthase

Query Match 99.9%; Score 1634; DB 2; Length 311;
Best Local Similarity 99.7%; Pred. No. 4.7e-139;
Matches 310; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDTVPFSLAELOQGHODEFRRLRDKGLFYLTDGCLTDTLKSADIVIDEFERGSE 60
Db 1 MDTVPFSLAELOQGHODEFRRLRDKGLFYLTDGCLTDTLKSADIVIDEFERGSE 60
QY 61 AEKAAVTSVPVPMRGFTGLESESTAOITNTGSDYSCMCSMGTAADNLPSPGDFERIMT 120
Db 61 AEKAAVTSVPVPMRGFTGLESESTAOITNTGSDYSCMCSMGTAADNLPSPGDFERIMT 120
QY 121 QYFDROYTASRAVAREVLARATGTETPDGVEAFDLCEPILRFRYPPOVPEHRSABEOPLRM 180
Db 121 QYFPROYTASRAVAREVLARATGTETPDGVEAFDLCEPILRFRYPPOVPEHRSABEOPLRM 180
QY 181 APHYDLSMVTLLIQOTPCANGFVSLQAEVGAFTDLPRPDAVLFVCGAIALVTGQGVKA 240
Db 181 APHYDLSMVTLLIQOTPCANGFVSLQAEVGAFTDLPRPDAVLFVCGAIALVTGQGVKA 240
QY 241 PRHHVAAPRRDQIAGSSRTSSVFELRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
Db 241 PRHHVAAPRRDQIAGSSRTSSVFELRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
QY 301 NYVNIRRTSKA 311
Db 301 NYVNIRRTSKA 311

RESULT 3

S30900
deacetoxycephalosporin C synthetase - Streptomyces lactamdurans
C;Species: Streptomyces lactamdurans
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C;Accession: S30900; S28383
R;Coque, J.J.R.; Martin, J.F.; Liras, P.
Mol. Gen. Genet. 236, 453-458, 1993
A>Title: Characterization and expression in Streptomyces lividans of cefD and cefE genes
clavuligerus.
A;Reference number: S30900; MUID:93173127; PMID:8437592
A;Accession: S30900
A;Molecule type: DNA
A;Residues: 1-314 <COO>
A;Cross-references: UNIPROT:O03047; EMBL:Z13974; NID:949299; PIDD:CAA76376.1; PID:949300
A;Note: the source is designated as Nocardia lactamdurans
C;Genetics:
A;Gene: cefE
C;Superfamily: isopenicillin N synthase

Query Match 73.1%; Score 1195.5; DB 2; Length 314;
Best Local Similarity 70.4%; Pred. No. 1.2e-99;
Matches 219; Conservative 40; Mismatches 51; Indels 1; Gaps 1;

QY 2 DTTVPFSLAELOQGHODEFRRLRDKGLFYLTDGCLTDTLKSADIVIDEFERGSE 60
Db 3 DATVPFDLAEIRGLHQEERHCLREKGVYLGITGLPABADHAGRELAVIDFDHGTGTE 62
QY 61 AEKAAVTSVPVPMRGFTGLESESTAOITNTGSDYSCMCSMGTAADNLPSPGDFERIMT 120
Db 63 AEKAAVTSVPVPMRGFTGLESESTAOITNTGSDYSCMCSMGTAADNLPSPGDFERIMT 120

QY 121 QYFDROYTASRAVAREVLARATGTETPDGVEAFDLCEPILRFRYPPOVPEHRSABEOPLRM 180
Db 123 DYFAARTRASQDVARQVLSVGAPEVGMADFDCPLRLKRFVPEPDRVAEBOPLRM 182
QY 181 APHYDLSMVTLLIQOTPCANGFVSLQAEVGAFTDLPRPDAVLFVCGAIALVTGQGVKA 240
Db 183 APHYDLSIVTLIHQTCANGFVSLQAEVGVSYVDIPAGCAVLFVCGAIALVTGQGVKA 242
QY 241 PRHHVAAPRRDQIAGSSRTSSVFELRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
Db 243 PRHHVAAPRRDQIAGSSRTSSVFELRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 302
QY 301 NYVNIRRTSKA 311
Db 303 NYVNIRRTSKA 313

RESULT 4

S40253
deacetoxycephalosporin C synthetase - Streptomyces lactamdurans
C;Species: Streptomyces lactamdurans
C;Date: 19-May-1994 #sequence_revision 26-May-1995 #text_change 09-Jul-2004
C;Accession: S40253
R;Coque, J.
submitted to the EMBL Data Library, February 1993
A;Reference number: S40253
A;Accession: S40253
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-310 <COO>
A;Cross-references: UNIPROT:P42219; EMBL:Z21687; NID:9438193; PIDD:CAA79803.1; PID:943819;
A;Note: the source is designated as Nocardia lactamdurans
C;Superfamily: 1-aminocyclopropane-1-carboxylate oxidase

Query Match 59.4%; Score 972.5; DB 2; Length 310;
Best Local Similarity 61.5%; Pred. No. 1.3e-79;
Matches 187; Conservative 37; Mismatches 79; Indels 1; Gaps 1;

QY 2 DTTVPFSLAELOQGHODEFRRLRDKGLFYLTDGCLTDTLKSADIVIDEFERGSA 61
Db 3 DKTVPFSLAELOQGHODEFRRLRDKGLFYLTDGCLTDTLKSADIVIDEFERGSA 61
QY 62 EKAAVTSVPVPMRGFTGLESESTAOITNTGSDYSCMCSMGTAADNLPSPGDFERIMTQ 121
Db 62 EKAAVTSVPVPMRGFTGLESESTAOITNTGSDYSCMCSMGTAADNLPSPGDFERIMTQ 121
QY 122 YFDROYTASRAVAREVLARATGTETPDGVEAFDLCEPILRFRYPPOVPEHRSABEOPLRM 181
Db 122 YFDROYTASRAVAREVLARATGTETPDGVEAFDLCEPILRFRYPPOVPEHRSABEOPLRM 181
QY 182 APHYDLSMVTLLIQOTPCANGFVSLQAEVGAFTDLPRPDAVLFVCGAIALVTGQGVKA 241
Db 182 APHYDLSMVTLLIQOTPCANGFVSLQAEVGAFTDLPRPDAVLFVCGAIALVTGQGVKA 241
QY 242 RHHVAAPRRDQIAGSSRTSSVFELRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 301
Db 242 RHHVAAPRRDQIAGSSRTSSVFELRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 301
QY 302 YVNI 305
Db 302 YVNI 305

RESULT 5

A39204
deacetoxycephalosporin C synthetase hydroxylase - Streptomyces clavuligerus
C;Species: Streptomyces clavuligerus
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C;Accession: A39204; A23713
R;Kovacevic, S.; Miller, J.R.
J. Bacteriol. 173, 398-400, 1991
A>Title: Cloning and sequencing of the beta-lactam hydroxylase gene (cefE) from Streptomyces

A:Reference number: A39204; MUID:91100311; PMID:1987130
A:Molecule type: DNA
A:Residues: 1-318 <KOV>
A:Cross-references: UNIPROT:P42220; GB:M63809; GB:M37186; NID:g153206; PIDN:AAA26716.1;
R:Baker, B.J.; Doolittle, J.E.; Yeh, W.K.
J. Biol. Chem. 266, 5087-5093, 1991
A>Title: Deacetoxycephalosporin C hydroxylase of Streptomyces clavuligerus. Purification
A:Reference number: A23713; MUID:91161600; PMID:2002049
A:Accession: A23713
A:Molecule type: protein
A:Residues: 2-29; 92-100 <BAK>
C:Genetics:
A:Gene: celf
C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase

Query Match 57.9%; Score 946.5; DB 2; Length 318;
Best Local Similarity 59.0%; Pred. No. 2.8e-77;
Matches 181; Conservative 43; Mismatches 80; Indels 3; Gaps 1;

Dy 2 DTTVPFSLAEIQQGLHODEFRRCIRDKGLFYLTDCGLTDTLKSAXDIVDFEHSSEA 61
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Dp 3 DTPVPIFLIALREGADDEKRECYTGNGVFYLLGYAGDGHDHLATDANDFPANTGEA 62
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 62 EKRAVTSVPMRRKGFTGSESTAOITNTGSYSDSMCSMGADNLFPSGDFERITWQ 121
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 63 EKAIVTDVPMRRGYSALAEASTAOVRTGSYTSDMSFSMGISGNVFPSPFERRWTE 122
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 122 YFDQVTSRAVARREVLRATG---TERPDGVEAFLEDCPLLRPRYPPOVPEHRSAEQPL 178
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 123 YFDCLYAQAQETAKLVTLASGGDAELVGSIDELLDDAPVLRLTFPEVPEHRSAEHPR 182
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 179 RMAPHYDLISMTLLIQOTPCANGFVSLOAEVGAAFTDPYRPDAVLVEFGAIATLVGTGOV 238
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 183 RMAHYDLSITLFHQPTCANGFVSLQAEIGELVSLPVVEDAVVMCGAMAPLATGAL 242
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 239 KAPRHVAAPERRODIAGSSRTSSVFEKLRENADFTFSVPALARCEGPDVSLDETATPDWI 298
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 243 PAPRHVRSPAGMRGSGDRSTSVFFLPRTDFSFSAKARSYGILAVLDMETATPDWI 302
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 299 GGNVNI 305
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 303 GTNYVTM 309
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 6
S54100
deacetoxycephalosporin C synthetase 1 - Lysobacter lactamgens (strain YK30)
C:Species: Lysobacter lactamgens
A:Variety: strain YK30
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
R:Kimura, H.; Iizawa, M.; Miyashita, H.; Shimizu, Y.; Sumino, Y.; Suzuki, M.
Submitted to the EMBL Data Library, October 1990
A:Description: Gene cluster involved in the cephalosporin biosynthesis from Lysobacter 1
A:Reference number: S54099
A:Accession: S54100
A:Molecule type: DNA
A:Residues: 1-319 <KIT>
A:Cross-references: UNIPROT:Q48740; EMBL:X56660; NID:g769806; PIDN:CAA39984.1; PID:g7698
A:Experimental source: strain YK30
C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase
C:Keywords: antibiotic biosynthesis

Query Match 57.7%; Score 944; DB 2; Length 319;
Best Local Similarity 58.4%; Pred. No. 4.8e-77;
Matches 178; Conservative 45; Mismatches 82; Indels 0; Gaps 0;

Oy 2 DTTVPFSLAEIQQGLHODEFRRCIRDKGLFYLTDCGLTDTLKSAXDIVDFEHSSEA 61
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 3 DSGQIFPLDELHGVRILDSFRKSLPERGVTVYAREDSIKTEHAKAMDADVMDLFPENGSAE 62
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 62 EKRAVTSVPMRRKGFTGSESTAOITNTGSYSDSMCSMGADNLFPSGDFERITWQ 121
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Db 63 QKQALNNLTPNVRKGSDELEASTARTYKGEETDYSKVNISIGLTDLPSPAPFAELMTWG 122

Qy 122 YEDROYTASRAAAREWELRATGTETPDGVEAFDLDCBELLRRYEPQVDEHRSABEQPLRMA 181

Db 123 YDFRFYARQDILARSLVLRASDAGLDBESVDFELDCDPLLRFRFPPEVEDRCABEPPRMA 182

Qy 182 PHYDLSMTWLLIOOTPCANGFVSIQAEVGAFTDLPYRPDAVLVFCGAIATLVTCGQYKAP 241

Db 183 PHYDLSMTWLLIOOTPCANGFVSIQAEVGAFTDLPYRPDAVLVFCGAIATLVTCGQYKAP 242

Qy 242 RHVVAAPRRDQJAGSRTSSVFFELRPNADFTFESVPLAREGFPVSLDGETATPDWIGN 301

Db 243 RHQVAPSPMQORVGSRTSSVFFELRPKPPDFRFSVPLARASGLDVDFGTATATFEGEWIGN 302

Qy 302 YYNIR 306

Db 303 YYNLR 307

RESULT 7

A29711

deaceoxycephalosporin C synthetase - fungus (Acremonium sp.)

N:Alternate names: expandase; hydroxylase

C:Species: Acremonium sp.

C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004

C:Accession: A29711; A41864

R:Samson, S.M.; Doerzaf, J.B.; Slisz, M.L.; Becker, G.W.; Van Frank, R.M.; Veal, L.E.; Y

Rio/Technology 5, 1207-1214, 1987

A:Title: Cloning and expression of the fungal expandase/hydroxylase gene involved in cep

A:Reference number: A29711

A:Accession: A29711

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-332 <SAM>

A:Cross-References: UNIPROT:P11935

A:Note: the source is designated as Cephalosporium acremonium

R:Gutierrez, S.; Velasco, J.; Fernandez, F.J.; Martin, J.F.

J. Bacteriol. 174, 3056-3064, 1992

A:Title: The cefG gene of Cephalosporium acremonium is linked to the cefEF gene and enco

A:Reference number: A41864; MUID:92234966; PMID:1569032

A:Accession: A41864

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-76, 'L' <GUT>

A:Experimental source: strain C10

A:Note: sequence extracted from NCBI backbone (NCBIN:104773, NCBI:P.97574); this ORF is m

A:Note: the source is designated as Cephalosporium acremonium

C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase

Query Match 57.1%; Score 933.5; DB 2; Length 332;

Best Local Similarity 57.1%; Pred. No. 4, 4e-76;

Matches 178; Conservative 43; Mismatches 90; Indels 1; Gaps 1;

Qy 1 MDTVPTFSLAEIQGLHODEFRRCILDKGLPYLTDGCLTDTLKSAXDIVDFFEKGS 60

Db 1 MTSKVPVFRILDLKSGKVLTELBAVTTKGIFLTSGVLDDHDTASARECVDPFKNGSE 60

Qy 61 AEKRAVTSVPTMRGFTGLSEBSTAQITNTGSSYSIMCYSNGTADNLPSPGDFEKIWT 120

Db 61 EKKRAVTTADRNARRGFSALWEVESTAVVETGKSYSTCTYSMGIGNLPPNGGFEDVMQ 120

Qy 121 QYFDQYATASRAVAREVLRATGTETPDG-GVEAFDLDCBELLRRYEPQVDEHRSABEQPLR 179

Db 121 DYFDQYATASRAVAREVLRATGTETPDG-GVEAFDLDCBELLRRYEPQVDEHRSABEQPLR 180

Qy 180 MAPHYDLSMTWLLIOOTPCANGFVSIQAEVGAFTDLPYRPDAVLVFCGAIATLVTCGQYK 239

Db 181 MAPHYDLSMTWLLIOOTPCANGFVSIQAEVGAFTDLPYRPDAVLVFCGAIATLVTCGQYK 240

Qy 240 AFRHVAAPRRDQJAGSRTSSVFFELRPNADFTFESVPLAREGCFDVSLDGETATPDWIG 299

Db 241 AFRHVAAPRRDQJAGSRTSSVFFELRPNADFTFESVPLAREGCFDVSLDGETATPDWIG 300

```
QY      300 GNYVNIRRTSKA 311
          ||||:|
Db      301 GNYVMRRDKPA 312
```

RESULT 8

deactoxyccephalosporin C synthetase 2 - *lysobacter lactamgens* (strain YK90)
C/Species: *Lysobacter lactamgens*
A/Variety: strain YK90
C/Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C/Accession: S54101
R/Kimura, H.; Izaa, M.; Miyashita, H.; Shimizu, Y.; Sumino, Y.; Suzuki, M.
submitted to the EMBL Data Library, October 1990
A/Description: Gene cluster involved in the cephalosporin biosynthesis from *lysobacter* 1
A/Reference number: S54099
A/Accession: S54101
A/Molecule type: DNA
A/Residues: 1-313 <KIM>
A/Cross-references: UNIPROT:Q04874; EMBL:X56660; NID:g769806; PIDN:CAA39985.1; PID:g769806
A/Experimental source: strain YK90
C/Superfamily: l-aminocyclopropane-1-carboxylate oxidase
/Keywords: antibiotic biosynthesis

Query Match	52.5%;	Score 859;	DB 2;	Length 313;
Best Local Similarity	53.2%;	Pred. No. 2e-69;		
Matches 166;	Conservative 48;	Mismatches 90;	Indels 8;	Gaps 2;

[illegible]

RESULT 9

isopenicillin N synthase (EC 1.14.11.-) pcbc [similarity] - Streptomyces griseus (scral)
C:Species: Streptomyces griseus
C:Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 09-Jul-2004
C:Accession: A6115
R:Garcia-Dominguez, M.; Liras, P.; Martin, J.F.
Antimicrob. Agents Chemother. 35, 44-52, 1991
A:Title: Cloning and characterization of the isopenicillin N synthase gene of Streptomyces
avilignus.
A:Reference number: A61155; MUID:91197089; PMID:1901702
A:Accession: A61155
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-329 <GAR>
A:Cross-references: UNIPROT:Q54243; GB:X54609; NID:G509097; PIDN:CA38431.1; PID:G509098
C:Genetic8:
A:Gene: pcbc
A:Superfamily: isopenicillin N synthase

C:Keywords: antibiotic biosynthesis; ascorbic acid; iron; metalloprotein; oxidoreductase
F148,212,268/Binding site: iron (His) #status predicted

Query Match	10.9%;	Score 178;	DB 2;	Length 329;
Best Local Similarity	24.2%;	Pred. No. 3.3e-08;		
Matches 72;	Conservative 40;	Mismatches 126;	Indels 60;	Gaps 12;

```

QY      5 VPTSLAELOOGQHOD-----EPRCLRDGGLFVLPDGGTLDEKAKDVIDFPEHG  58
Db      10 VPTDISLSSGDDADDKKVAQELINKACREGFFYASHHG--DVGLLKVAVEFHNTM  66
QY      59 SEAEK----RAVTSVPVPTMRGFTGLESESTAQITNTGYSYDSMCY-----  101
Db      67 TDEEKYDLAIYAANNKNNPRTNGY-----YAAVKGKAAVESWCYNLSPFSFHDPI  117
QY      102 SMGT--ADNLFPBGD----PERIMTOYEDRQYTAASRAVAREVLRACTEEDBGVEA---  151
Db      118 RSGPTMEGNIWPEBKHQHQRPRPCEQYRRVFSLSKYLWGFALALD-KEDDFDALS  176
QY      152 ---FLDCEPLRLFRFYFPQVPEHRSABE-QPLRMAPHYDLSNVTLLIQOTPCNANGFVSLQAE  207
Db      177 LADTLNATLTIHYRLBEDYRPVKVGRDGETKLSFEDHLDVSNITVLFQREYQN---LQVE  232
QY      208 VGGAFTDLPYRPDAVLVFCGALATLVTGQGVKAPRHVAALARRROLDAGSSRTSSVFL  265
Db      233 TADWMODLPYTSGENEFLVNVCGTYMGVLTINDYRPPAHRY-----KFINAEKLSLPEFL  284

```

RESULT 10

isopenicillin N synthase (EC 1.14.11.-) [similarity] - Streptomyces clavuligerus
C|Species: Streptomyces clavuligerus
C|Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 09-Jul-2004
C|Accession: A29894
R|Elekliw, B.K.; Aharonowitz, Y.; Mevarech, M.; Wolfe, S.; Vining, L.C.; Westlake, D.W.S.
Gene 62, 187-196, 1988
A|Title: Cloning and nucleotide sequence determination of the isopenicillin N synthetase
A|Reference number: A29894; MUID:88212175; PMID:3130293
A|Accession: A29894
A|Molecule type: DNA
A|Residues: 1-329 <LESS>
A|Cross-references: UNIPROT:P10621; GB:M19421; NID:g153329; PIDN:AAA26770.1; PID:g153329
C|Superfamily: isopenicillin N synthase
C|Keywords: antibiotic biosynthesis; ascorbic acid; iron; metalloprotein; oxidoreductase
E|48,212,268/Binding sites: iron (His) #status predicted

Query Match 10.3%; Score 168; DB 2; Length 329;
 Best Local Similarity 23.8%; Pred. No. 2.6e-07;
 Matches 77; Conservative 50; Mismatches 150; Indels 46; Gaps 14

QY 5 VPTSSLAEL-----QGGHODFERRCIRDGFLYLLDCCGLTDPELKSAKOIVIDFEHG 58

Db 10 VPTIDISPLFTDAAKKRAVAEEIHGACRGGSGFFATNHGV---DVQOQLOVNEFHGAM 66

QY 59 SEAEK-----RAVNSVPVTRGRGF-----TGLESSSTAQITVTGSYSYDSMCYSMTG---A 106

Db 67 TDQEHGDLIAHAYNPDPNPHVNGYYKAVPKKAABSFYLLNPDGDEGHPM-IAACTPVNHE 125

QY 107 DNLFPSSGD----FRIWTQYDFDROYTASRAVAREVLRATGTEPDGVAA-----FLDCE 156

Db 126 VNLMPPDERHRHFRFPCCGCGYRQWLKUSTVLMRGALALAG-RPEHFPPAALAEQDSLSV 184

QY 157 PLLRFRYFPQVBEHRSAAE-QPLMRAPHYDISMTWLLIQOTPCANGFVSLQAEVGCAFTDL 215

Db 185 SLIRPYLAEYBPVKTGPDGQLLSFEHDLDVSMITVLPTQVON---LQVETVDGMRDI 240

QY 216 PYRPDAVLVFCGALATLYTGQGVAKPRKHVAAAPRRDQIAGSSRRSSVFELAPNADFTFSV 275

Db 241 PTSENDFLVNGCTVMAHTVNDYFPAPNHRV-----KPVNAERLSPFLFNGCHEAVIE- 293

QY 276 PLARECGFDVSLDGETAFPODMI 298

Db 294 PVVPE-GASEEVRNEALSYGDYL 315

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.; Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: C64713
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-358 <STO>
A:Cross-references: UNIPROT:O80850; GB:AE002093; NID:g3201632; PIDN:AC20718.1; GSPDB:GN
C:Genetics:
A:Gene: At2g30830
A:Map position: 2
C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase

Query Match 8.8%; Score 144.5; DB 2; Length 358;
Best Local Similarity 22.3%; Pred. No. 3.7e-05;
Matches 60; Conservative 47; Mismatches 116; Indels 41; Gaps 9;

QY 23 RCLRDKGLFYLTCGLTDTLTKSAKDVIDFFHSGEAEKRAVTSPPVPMRGRFTGLSES 82
DB 78 RAAVEKEGFGFQVNHGIPLEVEESMKDGRGFHEDSEVKTFYSRDI-----T 126
QY 83 ESTAQITNTGSDYSYDMCY----SMGTADNLFPSGDF---ERIMTQYFDRQYASRAVA 134
DB 127 KKVKYNTNFDYSSQAAMNRDTLTWMAAPDVPAQADLPVICREIMLEYSKMMKLGELIF 186
QY 135 REVLRATGTEPDGGEAEFLDCEPLIRF--RYFPQVPEHRSAEQPLRNAPHYDLSWYTLI 192
DB 187 ELTSEALGLKPNHKE--LNCAKSISLSHYPPCP-----PDRFTGISSTHDISFITIL 240
QY 193 QQTCCANGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVTGGQVKAAPHNVAAARRDQ 252
DB 241 LQ---DHIGGLQVLHDGWDVPPNPBALIVNLGDLQLITNDKFWSEHRAVLNRGE 296
QY 253 IAGSSRTSSVFPLRPNADFTFSVP 276
DB 297 ----PRISSASF-----FMHTIP 310

RESULT 15
S33510
flavonol synthase - garden petunia
C:Species: Petunia x hybrida (garden petunia)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C:Accession: S33510
R:Holtan, T.A.
submitted to the EMBL Data Library, April 1993
A:Description: Cloning and expression of flavonol synthase from Petunia hybrida.
A:Reference number: S33510
A:Accession: S33510
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-348 <HOL>
A:Cross-references: UNIPROT:Q07512; EMBL:Z22543; NID:g311657; PIDN:CAA80264.1; PID:g3116
C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase
C:Keywords: iron; metalloprotein
F:87,234,290/Binding site: Iron (His) #status predicted

Query Match 8.7%; Score 143; DB 2; Length 348;
Best Local Similarity 22.3%; Pred. No. 4.9e-05;
Matches 59; Conservative 43; Mismatches 114; Indels 48; Gaps 9;

QY 27 RDKGLFYLTDCGLDTLTKSAKDIYIDFFHSGEAEKRAVT-SPVPTMRGRF-TGLSSES 84
DB 77 KEMGIFOLINHGIDPDEALADLQVKEFEFHVPOBEKELIAKTPGSNDIBGYSTLQKEV 136
QY 85 TAOITNTGSDYSYDMCYSMGTADNLF-----PSGDFERIY-----TOYFDRQYT 128
DB 137 EGR-----KGVNDHLFKIKIPPSAVNRYWPKNPSPSYREANEYGYKMRRE 181
QY 129 ASRAVAREVLATGTEPDGGEAEFLDCE--PLRFRYPQVPEHRSAEQPLRNAPHYDL 186

DB 182 VNDRIKFSLSLGLGHEHMEIAAGDEIVLLKINYPPPCPR----PDLAGVVAHTDM 237
QY 187 SWYTLIQQTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVTGGQVKAAPHNVAA 246
DB 238 SYITIL-----VPEVVOGLQVFKDGHWDVXYIPNALIVHIGQVEILSNKYSVYHRTT 293
QY 247 APRRDQIAGSSRTSSVFPLRPND 270
DB 294 VNK-----DKTRMSWVFLEPPSE 312

Search completed: September 19, 2005, 15:18:50
Job time : 32 secs

Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein, Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2005, 15:01:52; Search time 90 Seconds

(without alignments)
1769.518 Million cell updates/sec

Title: 10719236-1_232-1164

Perfect score: 1636

Sequence: 1 MDTWPTFSLAEIQGLHQD.....ATFQDMIGNVYNIKRTSKA 311

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1634	99.9	311	1	CEFE_STRCL
2	1385	84.7	311	1	Q7MYW9
3	1378	84.2	311	2	Q93FPA
4	1300	79.5	311	2	Q7MYW8
5	1195.5	73.1	314	1	CEFE_NOCUA
6	1012	61.9	312	2	Q7MYW7
7	972.5	59.4	310	1	CEFE_NOCUA
8	946.5	57.9	318	1	CEFE_STRCL
9	944	57.7	319	2	Q48740
10	933.5	57.1	332	1	EXSA_CEPAC
11	924.5	56.5	332	2	Q9P4T5
12	859	52.5	313	2	Q48741
13	427	26.1	82	2	Q7B3R2
14	183	11.2	329	1	IPNS_STRJU
15	180	11.0	327	2	Q6MCL3
16	178	10.9	329	1	IPNS_STRGR
17	169	10.3	366	2	Q987V6
18	168	10.3	329	1	IPNS_STRCL
19	168	10.3	348	2	Q6N7A0
20	166.5	10.2	337	2	Q6N1B6
21	165	10.1	329	2	Q9ANU0
22	164	10.0	326	1	IPNS_PLASS
23	164	10.0	326	2	IPNS_LYSIA
24	163	10.0	262	2	Q9PAC2
25	161	9.8	342	2	Q83YD9
26	160	9.8	262	2	Q9PAC0
27	158	9.7	262	2	Q9PAB6
28	158	9.7	262	2	Q9PAC1
29	156.5	9.6	403	2	Q8M2G6
30	156.5	9.6	403	2	Q9VCZ2
31	156	9.5	262	2	Q9PAC3

32	155.5	9.5	408	2	Q614Y3	Q614Y3 oryza sativ
33	155	9.5	262	2	Q9FAB8	Q9FAB8 streptomyc
34	155	9.5	262	2	Q9FAB9	Q9FAB9 streptomyc
35	155	9.5	331	2	Q91CZ2	Q91CZ2 streptomyc
36	154.5	9.4	366	2	Q942Y0	Q942Y0 oryza sativ
37	152	9.3	321	2	Q889J1	Q889J1 pseudomonas
38	147	9.0	333	1	IPNS_STRLP	IPNS_STRLP streptomyc
39	146.5	9.0	348	2	Q8S8B9	Q8S8B9 oryza sativ
40	146.5	9.0	348	2	Q7XC81	Q7XC81 oryza sativ
41	145	8.9	362	2	Q7S1P9	Q7S1P9 oryza sativ
42	145	8.9	367	2	Q76129	Q76129 ipomoea nil
43	144.5	8.8	358	2	Q80850	Q80850 arabidopsis
44	144.5	8.8	389	2	Q8RVF5	Q8RVF5 oryza sativ
45	144.5	8.8	389	2	Q7EC22	Q7EC22 oryza sativ

ALIGNMENTS

```
RESULT 1
CEFE_STRCL STANDARD, PRT, 311 AA.
ID_CEFSTRCL
AC P18548;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Deacetoxycephalosporin C synthetase (EC 1.14.20.1) (DNOCS)
DE (Expandase).
GN Name=cefe;
OS Streptomyces clavuligerus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=1901;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27064 / DSM 738 / NRRL 3585;
RX MEDLINE=89123150; PubMed=2644235;
RA Kovacevic S., Weigel B.J., Tobin M.B., Ingolia T.D., Miller J.R.;
RT "Cloning, characterization, and expression in Escherichia coli of the
RT Streptomyces clavuligerus gene encoding deacetoxycephalosporin C
RT synthetase."
RL J. Bacteriol. 171:754-760(1989).
CC -! FUNCTION: Catalyzes the step from penicillin N to deacetoxy-
CC cephalosporin C.
CC -! CATALYTIC ACTIVITY: Penicillin N + 2-oxoglutarate + O(2) =
CC -! COFACTOR: Iron and ascorbate.
CC -! PATHWAY: Cephalosporin antibiotics biosynthesis.
CC -! SIMILARITY: Belongs to the iron/ascorbate-dependent oxidoreductase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M32324; AAA26715.1; -.
CC PIR: T52312; T52312.
CC PDB: 1DCS; X-ray; @=1-311.
CC PDB: 1ESU; X-ray; A=1-306.
CC PDB: 1HUF; X-ray; A=1-311.
CC PDB: 1HUG; X-ray; A=1-311.
CC PDB: 1RXF; X-ray; @=1-311.
CC PDB: 1RXG; X-ray; @=1-311.
CC PDB: 1UNB; X-ray; A=1-311.
CC PDB: 1UO9; X-ray; A=1-311.
CC PDB: 1UOB; X-ray; A=1-311.
CC PDB: 1UOF; X-ray; A=1-311.
CC PDB: 1UOG; X-ray; A=1-311.
```

```
DR InterPro; IPR005123; 2OG-Ferri_Oxy.
DR InterPro; IPR002057; Isopen_N_synth.
DR Pfam; PF03171; 2OG-Ferri_Oxy; 1.
DR PROSITE; PS00185; IPNS_1; FALSE_NEG.
DR PROSITE; PS00186; IPNS_2; 1.
KM 3D-structure; Antibiotic biosynthesis; Iron; Oxidoreductase;
KM Vitamin C.
FT STRAND 7 9
FT HELIX 10 14
FT TURN 15 16
FT HELIX 19 28
FT TURN 29 29
FT STRAND 31 35
FT HELIX 41 57
FT HELIX 60 65
FT TURN 66 66
FT STRAND 67 67
FT STRAND 76 79
FT STRAND 99 103
FT HELIX 113 140
FT TURN 141 142
FT TURN 146 147
FT HELIX 149 153
FT STRAND 158 164
FT STRAND 180 185
FT STRAND 189 195
FT TURN 198 199
FT STRAND 204 208
FT TURN 209 210
FT STRAND 211 214
FT TURN 219 220
FT STRAND 222 226
FT HELIX 228 233
FT TURN 234 235
FT STRAND 243 245
FT TURN 254 255
FT STRAND 258 265
FT TURN 269 270
FT STRAND 272 274
FT HELIX 275 280
FT TURN 281 282
FT STRAND 291 293
FT HELIX 294 298
FT STRAND 305 306
SQ SEQUENCE 311 AA; 3455 MW; 9C64E1FC37F524BC CRC64;

Query Match 99.9%; Score 1634; DB 1; Length 311;
Best Local Similarity 99.7%; Pred. No. 1.5e-139;
Matches 310; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDTVPFSLAEIQGLHODEFRCLRDKGLFYLTDGSLTDTLKSAXDIVIDFEEHGE 60
DB 1 MDTVPFSLAEIQGLHODEFRCLRDKGLFYLTDGSLTDTLKSAXDIVIDFEEHGE 60
QY AEKRAVTSVPVPTMRGFTGLSESTAOITNTGSYSIDSMCYSMGTADNLFPSCGFERRIWT 120
DB AEKRAVTSVPVPTMRGFTGLSESTAOITNTGSYSIDSMCYSMGTADNLFPSCGFERRIWT 120
QY AEKRAVTSVPVPTMRGFTGLSESTAOITNTGSYSIDSMCYSMGTADNLFPSCGFERRIWT 120
DB AEKRAVTSVPVPTMRGFTGLSESTAOITNTGSYSIDSMCYSMGTADNLFPSCGFERRIWT 120
QY 121 QYFDROYTASRAVAEVLATGTEPDGVEAFLDCEPLLRFYFPQVPEHRSABEQPLRM 180
DB 121 QYFDROYTASRAVAEVLATGTEPDGVEAFLDCEPLLRFYFPQVPEHRSABEQPLRM 180
QY 121 QYFDROYTASRAVAEVLATGTEPDGVEAFLDCEPLLRFYFPQVPEHRSABEQPLRM 180
DB 121 QYFDROYTASRAVAEVLATGTEPDGVEAFLDCEPLLRFYFPQVPEHRSABEQPLRM 180
QY 181 APHYDLSMWTLIQOTPCANGFVSLQAEVGAFTDLPRPDAVLVFCGAIALTLVTGGQVKA 240
DB 181 APHYDLSMWTLIQOTPCANGFVSLQAEVGAFTDLPRPDAVLVFCGAIALTLVTGGQVKA 240
QY 241 PRHVAAPRRDOJAGSSRTSSVFFLRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
DB 241 PRHVAAPRRDOJAGSSRTSSVFFLRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
QY 301 NYVNIIRRTSKA 311
DB 301 NYVNIIRRTSKA 311
```

```
RESULT 2
ID Q7MYW9 PRELIMINARY; PRT; 311 AA.
AC Q7MYW9;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Deacetoxycephalosporin C synthase.
OS Streptomyces ambiofaciens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1899;
RN [1]
RP SEQUENCE FROM N.A.
RA Heu J.S., Yang Y.B., Wei C.L., Tsai Y.C.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the iron/ascorbate-dependent oxidoreductase
   family.
DR EMBL; AY18742; AAP85380.1; -.
DR HSSP; P18548; IE51.
DR GO; GO:0016021; C:Integral to membrane; IEA.
DR GO; GO:0019867; C:outer membrane; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0015288; F:porin activity; IEA.
DR InterPro; IPR005123; 2OG-Ferri_Oxy.
DR InterPro; IPR002057; Isopen_N_synth.
DR Pfam; PF03171; 2OG-Ferri_Oxy; 1.
DR PROSITE; PS00186; IPNS_2; UNKNOWN_1.
KM Iron; Oxidoreductase; Forin.
SQ SEQUENCE 311 AA; 34396 MW; 190869D7124B96C4 CRC64;

Query Match 84.7%; Score 1385; DB 2; Length 311;
Best Local Similarity 81.4%; Pred. No. 5.5e-117;
Matches 253; Conservative 33; Mismatches 25; Indels 0; Gaps 0;

QY 1 MDTVPFSLAEIQGLHODEFRCLRDKGLFYLTDGSLTDTLKSAXDIVIDFEEHGE 60
DB 1 MDTVPFSLAEIQGLHODEFRCLRDKGLFYLTDGSLTDTLKSAXDIVIDFEEHGE 60
QY AEKRAVTSVPVPTMRGFTGLSESTAOITNTGSYSIDSMCYSMGTADNLFPSCGFERRIWT 120
DB AEKRAVTSVPVPTMRGFTGLSESTAOITNTGSYSIDSMCYSMGTADNLFPSCGFERRIWT 120
QY AEKRAVTSVPVPTMRGFTGLSESTAOITNTGSYSIDSMCYSMGTADNLFPSCGFERRIWT 120
DB AEKRAVTSVPVPTMRGFTGLSESTAOITNTGSYSIDSMCYSMGTADNLFPSCGFERRIWT 120
QY 121 QYFDROYTASRAVAEVLATGTEPDGVEAFLDCEPLLRFYFPQVPEHRSABEQPLRM 180
DB 121 QYFDROYTASRAVAEVLATGTEPDGVEAFLDCEPLLRFYFPQVPEHRSABEQPLRM 180
QY 121 QYFDROYTASRAVAEVLATGTEPDGVEAFLDCEPLLRFYFPQVPEHRSABEQPLRM 180
DB 121 QYFDROYTASRAVAEVLATGTEPDGVEAFLDCEPLLRFYFPQVPEHRSABEQPLRM 180
QY 181 APHYDLSMWTLIQOTPCANGFVSLQAEVGAFTDLPRPDAVLVFCGAIALTLVTGGQVKA 240
DB 181 APHYDLSMWTLIQOTPCANGFVSLQAEVGAFTDLPRPDAVLVFCGAIALTLVTGGQVKA 240
QY 241 PRHVAAPRRDOJAGSSRTSSVFFLRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
DB 241 PRHVAAPRRDOJAGSSRTSSVFFLRPNADFTFSVPLARECGFVSLDGETATFQDWIGG 300
QY 301 NYVNIIRRTSKA 311
DB 301 NYVNIIRRTSKA 311

RESULT 3
Q93FD4 PRELIMINARY; PRT; 311 AA.
ID Q93FD4;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Deacetoxycephalosporin C synthase.
OS Streptomyces jumoniensis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
```

```

OX NCBI_TaxId=1945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29864;
RA Sim T.S., Sim J.;
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
CC -! SIMILARITY: Belongs to the iron/ascorbate-dependent oxidoreductase
   family.
CC
DR EMBL: AF317908; AAL09460.1; -.
DR HSSP: P18548; 1E51.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0019867; C:outer membrane; IEA.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0015288; F:porin activity; IEA.
DR InterPro: IPR005123; 2OG-Ferri_Oxy.
DR Pfam: PF03171; 2OG-Ferri_Oxy; 1.
DR PROSITE: PS00186; IPNS 2; UNKNOWN_1.
DR Iron; Oxidoreductase; Porin.
SQ: SEQUENCE 311 AA; 34242 MW; B821ED796A4F5C1A CRC64;

Query Match      84.2%; Score 1378; DB 2; Length 311;
Best Local Similarity 81.7%; Pred. No. 2,4e-116;
Matches 254; Conservative 28; Mismatches 29; Indels 0; Gaps 0;

QY 1 MDTTVPFSLAELOQGLHODEFRRLDKGLFYLTDDCGITDTELKSAKDVIIDFFEHGSE 60
DB 1 MDTTVPFSLAELOQGLHODEFRRLDKGLFYLTDDCGITDTELKSAKDVAIDFFEHGTE 60
QY 61 AEKRAVSPVPTMRGFTGLSESTAOITNTGSYSDSMCMGTADNLFPSCGFERITWT 120
DB 61 EKKRAVTSITPTIRRGFTGLSESTAOITNTACTYSDSMCMGLADNVFPFGDFFERWT 120
QY 121 QYFDRQYASRAVAEVLRAATGTEPDGVEAFLLDCEPLLRFRYPQVPEHRSABEQPLRM 180
DB 121 HYFGHMYDVTQGEVARQVLTGTETEPYGVGESFVDCPEPLLRFRYPPEVEDRSABEQPLRM 180
QY 161 APHYDLSTVTLIQOTPCNGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVLTGGQYKA 240
DB 161 APHYDLSTVTLIQOTPCNGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVLTGGQYKA 240
QY 241 PRHHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSVPLARECGFVSLDGETATPDQWIGG 300
DB 241 PRHHVAAPRRDQIAGSSGRTSSVFPLRPDSDFSVSLARECGFDISLDGETATFDQWIEG 300
QY 301 NYVINIRRTSKA 311
DB 301 NYVINIRRTSEA 311

RESULT 4
Q7WYW8 PRELIMINARY; PRT; 311 AA.
ID Q7WYW8;
AC Q7WYW8;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Deacetoxycephalosporin C synthase.
OS Streptomyces chartreusii.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=1969;
RN [1]
RP SEQUENCE FROM N.A.
RA Hsu J.S., Yang Y.B., Wei C.L., Tsai Y.C.;
RL Submitted (JUN-2003) to the EMBL/Genbank/DBJ databases.
CC -! SIMILARITY: Belongs to the iron/ascorbate-dependent oxidoreductase
   family.
CC
DR EMBL: AY316743; AAP85381.1; -.
DR HSSP: P18548; 1E51.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0019867; C:outer membrane; IEA.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.

```

```

DR GO: GO:0015288; F:porin activity; IEA.
DR InterPro: IPR005123; 2OG-Ferri_Oxy.
DR InterPro: IPR002057; Isopen_N_synth.
DR Pfam: PF03171; 2OG-Ferri_Oxy; 1.
DR PROSITE: PS00186; IPNS 2; UNKNOWN_1.
DR Iron; Oxidoreductase; Porin.
SQ: SEQUENCE 311 AA; 34076 MW; C26880E90221B441 CRC64;

Query Match      79.5%; Score 1300; DB 2; Length 311;
Best Local Similarity 76.5%; Pred. No. 2,8e-109;
Matches 238; Conservative 32; Mismatches 41; Indels 0; Gaps 0;

QY 1 MDTTVPFSLAELOQGLHODEFRRLDKGLFYLTDDCGITDTELKSAKDVIIDFFEHGSE 60
DB 1 MDTTVPFSLAELOQGLHODEFRRLDKGLFYLTDDCGITDTELKSAKDVAIDFFEHGTT 60
QY 61 AEKRAVSPVPTMRGFTGLSESTAOITNTGSYSDSMCMGTADNLFPSCGFERITWT 120
DB 61 EKKRAVTSITPTIRRGFTGLSESTAOITNTACTYSDSMCMGLADNVFPFGDFFERWT 120
QY 121 QYFDRQYASRAVAEVLRAATGTEPDGVEAFLLDCEPLLRFRYPQVPEHRSABEQPLRM 180
DB 121 QYFDRQYASRAVAEVLRAATGTEPDGVEAFLLDCEPLLRFRYPQVPEHRSABEQPLRM 180
QY 161 APHYDLSTVTLIQOTPCNGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVLTGGQYKA 240
DB 161 APHYDLSTVTLIQOTPCNGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVLTGGQYKA 240
QY 241 PRHHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSVPLARECGFVSLDGETATPDQWIGG 300
DB 241 PRHHVAAPRRDQIAGSSRTSSVFPLRPNADFTFSVPLARECGFDISLDGETATFDQWIEG 300
QY 301 NYVINIRRTSKA 311
DB 301 NYVINIRRTSKA 311

RESULT 5
CEFE_NOCCLA STANDARD; PRT; 314 AA.
ID CEFE_NOCCLA
AC Q03047;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Deacetoxycephalosporin C synthetase (EC 1.14.20.1) (DAOS)
DB (Expandase).
GN Name=cefe;
OS Nocardia lactamdurans.
OC Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales;
OC Pseudonocardiaceae; Pseudonocardiaceae; Amycolatopsis.
OX NCBI_TaxId=1913;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93173127; PubMed=8437592;
RA Coque J.J.R., Martin J.F., Litras P.;
RT "Characterization and expression in Streptomyces lividans of cefD and
   cefE genes from Nocardia lactamdurans: the organization of the
   cephamycin gene cluster differs from that in Streptomyces
   clavuligerus."
RL Mol. Gen. Genet. 236:453-458 (1993).
CC -! FUNCTION: Catalyzes the step from penicillin N to deacetoxy-
   cephalosporin C.
CC -! CATALYTIC ACTIVITY: Penicillin N + 2-oxoglutarate + O(2) =
   deacetoxycephalosporin C + succinate + CO(2) + H(2)O.
CC -! COFACTOR: Iron and ascorbate.
CC -! PATHWAY: Cephalosporin antibiotics biosynthesis.
CC -! SIMILARITY: Belongs to the iron/ascorbate-dependent oxidoreductase
   family.
CC
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC -----
 DR EMBL/ Z13974; CAA78376.1; -
 DR PIR/ S30900; S30900.
 DR HSSP/ P18548; 1B51.
 DR InterPro/ IPR005123; 20G-FeII_Oxy.
 DR InterPro/ IPR02057; Isopen_N_synch.
 DR Pfam/ PF03171; 20G-FeII_Oxy; 1.
 DR PROSITE/ PS0185; IPNS_1; FALSE NEG.
 DR PROSITE/ PS0186; IPNS_2; FALSE NEG.
 KM Antibiotic biosynthesis; Iron; Oxidoreductase; Vitamin C.
 SQ SEQUENCE 314 AA; 34530 MW; DD07AD3D67A1BDF CRC64;

Query Match 73.1%; Score 1195.5; DB 1; Length 314;
 Best Local Similarity 70.4%; Pred. No. 8.3e-100;
 Matches 219; Conservative 40; Mismatches 51; Indels 1; Gaps 1;

QY 2 DTTVPFSLAELOQGHODEFRCLRDKGLFYLTDCGLTDTLKSADIVIDFEHGSBA 60
 DB 3 DATVPFSLAELOQGHODEFRCLRDKGLFYLTDCGLTDTLKSADIVIDFEHGSBA 62
 QY 61 AEKRAVTSVPVPMRGFTGLSESTAOITNTGYSYDSCYSGMTADNLPFGDFFERITQ 120
 DB 63 AEKRAVTPPTIRRGYAGLESESTAOITNTGKXTDYSMSYSGMTADNLPFGSAFEKAME 122
 QY 121 QYPRQYTAARAVREYLRTGTEPDGVEAFIDCEPLLRFPRYPQVPEHRSABEOLPMA 180
 DB 123 DYFAWRVTRASQDVARQVLTSGVAEPEVGMDFLDCEPLLRLRYPPEVEDRVAEEOPLEMA 182
 QY 181 APHYDLSMVTLLIOQTPCANGFVSLQAEVGAFTDLPYRPDAVIVFCGAIATLVYGGVKA 240
 DB 183 APHYDLSMVTLLIOQTPCANGFVSLQAEVGSYVDIPADGAVLVFCGAVATLVADGAIKA 242
 QY 241 PRHVAAPRRDQIAGSSRTSSVFPLRPVADFTFSPVPLARECGFVSLDGTATFQDMIGG 300
 DB 243 PKHVAAPGADKRGSSRTSSVFPLRPVADFTFSPVPLARECGFVSLDGTATFQDMIGG 302
 QY 301 NYNIRRTSKA 311
 DB 303 NYNIRRTSKA 313

RESULT 6

Q7WTW7 PRELIMINARY; PRT; 312 AA.
 AC Q7WTW7;
 DT 01-OCT-2003 (Tremblrel. 25, Created)
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
 DE Desacetylcephalosporin C synthase.
 OS Streptomyces sp. 65PH1.
 OC Bacteria; Actinobacteridae; Actinobacteriales;
 OC Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=237545;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=65PH1;
 RA Hsu J.S., Yang Y.B., Wei C.L., Tsai Y.C.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to the iron/ascorbate-dependent oxidoreductase
 CC family.
 DR EMBL/ AY318744; AAP85382.1; -
 DR HSSP/ P18548; 1B51.
 DR GO/ GO:0016021; C:Integral to membrane; IEA.
 DR GO/ GO:0019867; C:outer membrane; IEA.
 DR GO/ GO:0016491; F:oxidoreductase activity; IEA.
 DR GO/ GO:0015288; F:porin activity; IEA.
 DR InterPro/ IPR005123; 20G-FeII_Oxy;
 DR Pfam/ PF03171; 20G-FeII_Oxy; 1.
 KM Iron; Oxidoreductase; Porin.
 SQ SEQUENCE 312 AA; 34332 MW; 428DF7A15B802487 CRC64;

Query Match 61.9%; Score 1012; DB 2; Length 312;
 Best Local Similarity 60.3%; Pred. No. 3.5e-83;
 Matches 187; Conservative 47; Mismatches 76; Indels 0; Gaps 0;

QY 2 DTTVPFSLAELOQGHODEFRCLRDKGLFYLTDCGLTDTLKSADIVIDFEHGSBA 61
 DB 3 DTTVPFSLAELOQGHODEFRCLRDKGLFYLTDCGLTDTLKSADIVIDFEHGSBA 62
 QY 62 EKRVAVTSVPVPMRGFTGLSESTAOITNTGYSYDSCYSGMTADNLPFGDFFERITQ 121
 DB 63 EKRVAVTPPTIRRGYAGLESESTAOITNTGKXTDYSMSYSGMTADNLPFGSAFEKAME 122
 QY 122 YPRQYTAARAVREYLRTGTEPDGVEAFIDCEPLLRFPRYPQVPEHRSABEOLPMA 181
 DB 123 HFDNLSAARAVREYLRTGTEPDGVEAFIDCEPLLRFPRYPQVPEHRSABEOLPMA 182
 QY 182 PHYDLSMVTLLIOQTPCANGFVSLQAEVGAFTDLPYRPDAVIVFCGAIATLVYGGVKA 241
 DB 183 PHYDLSMVTLLIOQTPCANGFVSLQAEVGSYVDIPADGAVLVFCGAVATLVADGAIKA 242
 QY 242 RHHVAAPRRDQIAGSSRTSSVFPLRPVADFTFSPVPLARECGFVSLDGTATFQDMIGG 301
 DB 243 QHVRVAPGAGLRGSSRTSSVFPLRPVADFTFSPVPLARECGFVSLDGTATFQDMIGG 302
 QY 302 YVNIRRTSKA 311
 DB 303 YVNIRRTSKA 312

RESULT 7

CEFF_NOCIA STANDARD; PRT; 310 AA.
 ID CEFF_NOCIA
 AC P42219;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Desacetylcephalosporin C hydroxylase (EC 1.14.11.-)
 DE (Desacetylcephalosporin C synthetase) (DACS) (Beta-lactam hydroxylase).
 GN Name=ceff;
 OS Nocardia lactamdurans.
 OC Bacteria; Actinobacteridae; Actinomycetales;
 OC Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.
 OX NCBI_TaxID=1913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96287477; PubMed=8703431;
 RA Coque J., Enguita F.J., Cardoza R.E., Martin J.F., Liras P.;
 RT "Characterization of the ceff gene of Nocardia lactamdurans encoding a
 RT 3'-methylcephem hydroxylase different from the 7-cepem hydroxylase.";
 RL Appl. Microbiol. Biotechnol. 44:605-609(1996).
 CC -1- FUNCTION: Hydroxylation of desacetoxycephalosporin C in 3'position
 CC to form desacetylcephalosporin C.
 CC -1- PATHWAY: Cephalosporin biosynthesis.
 CC -1- SIMILARITY: Belongs to the iron/ascorbate-dependent oxidoreductase
 CC family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL/ Z21687; CAA79803.1; -
 DR PIR/ S40253; S40253.
 DR HSSP/ P18548; 1RXF.
 DR InterPro/ IPR005123; 20G-FeII_Oxy.
 DR Pfam/ PF03171; 20G-FeII_Oxy; 1.
 KM Antibiotic biosynthesis; Iron; Oxidoreductase; Vitamin C.
 SQ SEQUENCE 310 AA; 34364 MW; 9983EFLAC1B41E6 CRC64;

Query Match	Similarity	Score	DB	Length
Best Local Similarity	59.08%	Pred. No. 3,1e-77;		
Matches 181;	Conservative 43;	Mismatches 80;	Indels 3;	Gaps 1;
QY	2 DTTVPPTSFLAEIQQGLHODEFRRCRLRKGLFYLTDCCLTPTTELKSAKDVIDPFEHGESEA	61		
DB	3 DTPPIPIINLALARKGAQOEKRECVITGMGFFYLITGYAGSKDRLHLDITDAMDFFANGTEA	62		
QY	62 EKAAYVSPVPTMRRGFTGLESESTQAQITNTGYSYDYSKCYSMGTADNLPSGDFERITWQ	121		
DB	63 EKAAYITDVPFMRRGYALAESESTQAQYTRIGSYDYDSMSMGISGVFSPSEERVWTE	122		
QY	122 YFDQGYASRAVAEVLKATG---TEEDGGVEAFLODEBPLLRPFYFPOVGEHRSAAEQPL	176		
DB	123 YFDLCLVAAQOETAALVLTASGGYDAEIVGSLUDELDLADPVLRKLYFPEVPHRSAAHEPR	182		
QY	179 RMAPHYDLSMWTLTIQPTPCANGFVSLQAEVSGAFTDDPYRPDAVLVFCGAILATLVGQGV	236		
DB	183 RMAHYDLSITITFHQIFPCANGFVSLQAEIVGSELVSLPVVEDADVVMCGAMAPLATGAL	242		
QY	239 KAPPHVAAPRRDQIAGSSRTSSVFILRPNADFTFVSPLARECGFDVSLDGETATFQDWI	298		
DB	243 PAPPHHRSFGACMGREGSDRTSSVFILRPPTDSEFVSAAKARSYGLAVDLMETATFQDWI	302		
QY	299 GGNVYNI 305			
DB	303 GTNVYNI 309			
RESULT 9				
ID	Q48740	PRELIMINARY;	PRT;	319 AA.
AC	Q48740;			
DT	01-NOV-1996 (TrEMBLrel. 01, Created)			
DJ	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)			
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)			
OS	Deacetylcephalosporin C synthetase.			
DE	Lyso bacter lactamgenus.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;			
OC	Xanthomonadaceae; Lyso bacter.			
OX	NCBI_TaxID=39596;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=VK90;			
KA	MEDLINE=96287475; PubMed=8703429;			
RT	Kimura H., Izawa M., Sumino Y.,			
RT	"Molecular analysis of the gene cluster involved in cephalosporin			
RL	biosynthesis from lyso bacter lactamgenus VK90.";			
RL	Appl. Microbiol. Biotechnol. 44:589-596(1996).			
CC	-1- SIMILARITY: Belongs to the iron/ascorbate-dependent oxidoreductase			
CC	family.			
DR	EMBL; X55660; CA39984.1; -.			
DR	PIR; S54100; S54100.			
DR	HSSP; P18548; IRXF.			
DR	GO; GO:0016021; C:integral to membrane; IEA.			
DR	GO; GO:0019667; C:outer membrane; IEA.			
DR	GO; GO:0016481; F:oxidoreductase activity; IEA.			
DR	GO; GO:0015288; F:porin activity; IEA.			
DR	InterPro; IPR005123; 2OG-Fell_Oxy.			
DR	Pfam; PF03171; 2OG-Fell_Oxy. 1.			
KW	Iron; Oxidoreductase; Porin.			
SQ	SEQUENCE 319 AA; 35559 MW; 51040CD201B7C272 CRC64;			
Query Match				
Best Local Similarity	57.78%	Score 944;	DB 2;	Length 319;
Matches 178;	Conservative 45;	Mismatches 82;	Indels 0;	Gaps 0;
QY	2 DTTVPPTSFLAEIQQGLHODEFRRCRLRKGLFYLTDCCLTPTTELKSAKDVIDPFEHGESEA	61		
DB	3 DSGQITDDELDELHGRVLRDLSFKSLPERGVFYVRREDSITETIRAKAMDAVMDLFEUNGSAL	62		
QY	62 EKAAYVSPVPTMRRGFTGLESESTQAQITNTGYSYDYSKCYSMGTADNLPSGDFERITWQ	121		
DB	63 EKAAYITDVPFMRRGYALAESESTQAQYTRIGSYDYDSMSMGISGVFSPSEERVWTE	122		
QY	122 YFDQGYASRAVAEVLKATG---TEEDGGVEAFLODEBPLLRPFYFPOVGEHRSAAEQPL	176		
DB	123 YFDLCLVAAQOETAALVLTASGGYDAEIVGSLUDELDLADPVLRKLYFPEVPHRSAAHEPR	182		
QY	179 RMAPHYDLSMWTLTIQPTPCANGFVSLQAEVSGAFTDDPYRPDAVLVFCGAILATLVGQGV	236		
DB	183 RMAHYDLSITITFHQIFPCANGFVSLQAEIVGSELVSLPVVEDADVVMCGAMAPLATGAL	242		
QY	239 KAPPHVAAPRRDQIAGSSRTSSVFILRPNADFTFVSPLARECGFDVSLDGETATFQDWI	298		
DB	243 PAPPHHRSFGACMGREGSDRTSSVFILRPPTDSEFVSAAKARSYGLAVDLMETATFQDWI	302		
QY	299 GGNVYNI 305			
DB	303 GTNVYNI 309			

Db	1	MTSKVPEVPRLLDDLSGKVLTELAAVTTKGIFYLITESGLVDDBHTSARETCVDFPKNGSE	60
Oy	61	AERKAATVSPPVTMRRGTTGGTLESESTAOITNTGSISDYSMCYSMGTADNLPPSGDFEILWT	120
Db	61	EKKAAVTLAARNRARGFSALFEMESTAAVTEFGXKSDYSTCYSMGIIGNLFPNRGFEDVMQ	120
Oy	121	QYPRROYTASRAVREVLNATGTSPDG-GVEAFIDCEPLLRFRFPQVPEHRSAAEQPLR	179
Db	121	DYFRMTGAADVARAVINVSIGAPLAEBDIDDVECPDPLRLRFPEVPEBRVAEEBPLR	180
Oy	180	MAPHYDLISMVTLIOTPCANGFVSLQAEVGGAFTDLPYRPDAVLVFCGAIAATLVTSQOVK	239
Db	181	MGPHYDSTITLVHQTCANGFVSLQCEVGEJFDLPTLPAMVVFCGAVCTLATGKVX	240
Oy	240	APRRHVAAPRPDQIAGSSRTSSVFFLRPNADFTSVPLARECGDVSLDGETATFOWMIG	299
Db	241	APKRHVSPGHRDQGVSSRTSSVFELRPKDPSFNVOOSRWGMNVRI PSERTTFRMWLG	300
Oy	300	GNYVNIRTRSKA 311	
Db	301	GNYVMRRDKPA 312	
 RESULT 11 O9P4T5 PRELIMINARY; PRF: 332 AA.			
ID	O9P4T5	AC	O9P4T5
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DE	01-MAR-2004 (TREMBLrel. 26, Last annotation update)		
DR	Decetylcephalosporin C synthetase/hydroxylase.		
GN	Name=cefer;		
OS	Cephalosporium acremonium (Acromonium chrysogenum).		
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;		
CC	Hypocremomyces; Hypocreales; Hypocreaceae; mitosporic Hypocreaceae;		
OX	NCHI_TaxID=5044;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=A3/2;		
RA	Jekosch K., Nosek J., Kueck U.;		
RL	Submitted (JUL-2000) to the EMBL/GenBank/DDAJ databases.		
CC	-i- SIMILARITY: Belongs to the iron/ascorbate-dependent oxidoreductase family.		
CC	EMBL; AJ404737; CAB96750.1; -.		
DR	HSSP; P18548; IRXG.		
DR	GO; GO:0016021; C:integral to membrane; IEA.		
DR	GO; GO:0019867; C:outer membrane; IEA.		
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.		
DR	GO; GO:0015288; F:portin activity; IEA.		
DR	InterPro; IPR005123; 2OG-FeII_Oxy.		
DR	InterPro; IPR002057; Isopen_N_synth.		
DR	Pfam; PF03171; 2OG-FeII_Oxy; 1.		
DR	PROSITE; PS00185; IPNS_1; UNKNOWN_1.		
KW	PROSITE; PS00186; IPNS_2; UNKNOWN_1.		
KW	Iron; Oxidoreductase; Portin.		
SQ	SEQUENCE 332 AA; 36538 MW; E3DECESE9C28B79 CRC64;		
 Query Match 56.5%; Score 924.5; DB 2; Length 332; Best local similarity 56.7%; Pred. No. 3,2e-75; Matches 177; Conservative 43; Mismatches 91; Indels 1; Gaps 1;			
Oy	1	MDTVPTFSLAELOOGLHODEFRRCCLRDKGLFYLTDCGLTDTELKSAKDIVDPFHQSE	60
Db	1	MTSKVPVPRLLDDLSGKVLTELAAVTTKGIFYLTESGLVDDHTSARETCVDFPKNGSE	60
Oy	61	AERKAATVSPPVTMRRGTTGGTLESESTAOITNTGSISDYSMCYSMGTADNLPPSGDFEILWT	120
Db	61	EKKAAVTLAARNRARGFSALFEMESTAAVTEFGXKSDYSTCYSMGIIGNLFPNRGFEDVMQ	120
Oy	121	QYPRROYTASRAVREVLNATGTSPDG-GVEAFIDCEPLLRFRFPQVPEHRSAAEQPLR	179
Db	121	DYFRMTGAADVARAVINVSIGAPLAEBDIDDVECPDPLRLRFPEVPEBRVAEEBPLR	180
Oy	180	MAPHYDLISMVTLIOTPCANGFVSLQAEVGGAFTDLPYRPDAVLVFCGAIAATLVTSQOVK	239
Db	181	MGPHYDSTITLVHQTCANGFVSLQCEVGEJFDLPTLPAMVVFCGAVCTLATGKVX	240
Oy	240	APRRHVAAPRPDQIAGSSRTSSVFFLRPNADFTSVPLARECGDVSLDGETATFOWMIG	299
Db	241	APKRHVSPGHRDQGVSSRTSSVFELRPKDPSFNVOOSRWGMNVRI PSERTTFRMWLG	300
Oy	300	GNYVNIRTRSKA 311	
Db	301	GNYVMRRDKPA 312	


```

CC      cysteinyl-D-valine + O(2) = isopenicillin N + 2 H(2)O.
CC      -1- COFACTOR: Iron and ascorbate.
CC      -1- PATHWAY: Biosynthesis of penicillin and cephalosporin.
CC      -1- SIMILARITY: Belongs to the iron/ascorbate-dependent oxidoreductase
CC      family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC      or send an email to license@ebi.ac.uk).
CC      -----
DR      EMBL; M36687; AAA26772.1; -.
DR      HSSP; P05326; IODM.
DR      InterPro; IPR005123; 2OG-Ferri Oxy.
DR      InterPro; IPR002283; IPN synth.
DR      InterPro; IPR002057; Isopen N synth.
DR      Pfam; PF03171; 2OG-Ferri Oxy; 1.
DR      PRINTS; PR00682; IPNSYNTHASE.
DR      PROSITE; PS00185; IPNS_1; 1.
DR      PROSITE; PS00186; IPNS_2; 1.
DR      Antibiatic biosynthesis; Iron; Oxidoreductase; Vitamin C.
FT      METAL 212 212 Iron.
FT      METAL 214 214 Iron.
FT      METAL 268 268 Iron.
SQ      SEQUENCE 329 AA; 37305 MW; 9C8F1EB8FB8DF6C CRC64;

Query Match 11.2%; Score 183; DB 1; Length 329;
Best local Similarity 24.9%; Pred. No. 5e-08;
Matches 72; Conservative 45; Mismatches 130; Indels 42; Gaps 11;

QY 5 VPTFSL-----AELOQGHODEFRRLDKGLFYLTDCGLDTELSKADVIDPFEH 57
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 10 VPTIDISPLSGDDAKAKORVAQ-EINKARSGSGFYNSHGV---DVQLLDVNEFH 65

QY 58 GSEAEK-----RAVTSVPVTMRGFT---GLSESTAOITNTGSYSDYSGVSMGTAD- 107
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 66 MSDDKEDHLAINAYNKNDFPHRNGIYKAIKGGKAVESCYINPFSDDHPVITSETPMHE 125

QY 108 -NLFPSSGD---FRIWTQYFDRQYTSRAVAREVLATGTERPDGVAEFLDCE-----P 157
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 126 VNLMPDEKHPFRPFCEDDYRQLLRSTVIMRGVALAGRRDPFPEALAEADTLSSVS 185

QY 158 LLRRRYFPQVNEHRS-AEEOPLRMAPHYDLSMTLLIOQTPCANCFVSLQAEVGAFDLP 216
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 186 LIRPYLEEYPPVXTGADGTLSFEDHLDVSMITVLVQTEVQN-----LOVETVDGMODIP 241

QY 217 YRPDAVLVFCGAIATLVGGGVKAPRHVVAAPRRDQIAGSRTSSVFPL 265
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 242 RSDDFLVNCGTGMGHITHDYFPAPNHRV-----KFINERKLSLPFPL 284

RESULT 15
O6MQL3 PRELIMINARY; PRT; 327 AA.
AC O6MQL3;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Oxidoreductase.
GN OrderedLocustNames=Bd0452;
OS Bdellovibriio bacteriovorus.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Bdellovibrionales;
OC Bdellovibrionaceae; Bdellovibriio.
OX NCBI_TaxID=959;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD100 / DSM 50701 / ATCC 15356 / NCIB 9529;
RX PubMed=14752164; DOI=10.1126/science.1093027;
RA Rendulic S., Jagtap P., Rosinus A., Bpinger M., Baar C., Ianz C.,
  Keller H., Lambert C., Evans K.J., Goemann A., Meyer F.,

```

```

RA Sockett R.E., Schuster S.C.;
RT "A predator unmasked: life cycle of Bdellovibriio bacteriovorus from a
RT genomic perspective."
RL Science 303:689-692(2004).
CC      -1- SIMILARITY: Belongs to the iron/ascorbate-dependent oxidoreductase
CC      family.
CC      -----
DR      EMBL; BX842647; CAE78434.1; -.
DR      GO; GO:000506; F:iron ion binding; IEA.
DR      GO; GO:0016216; F:isopenicillin-N synthase activity; IEA.
DR      GO; GO:0017000; P:antibiotic biosynthesis; IEA.
DR      InterPro; IPR005123; 2OG-Ferri Oxy.
DR      InterPro; IPR002283; IPN synth.
DR      Pfam; PF03171; 2OG-Ferri Oxy; 1.
DR      PRINTS; PR00682; IPNSYNTHASE.
KM Complete proteome; Iron; Oxidoreductase.
SQ      SEQUENCE 327 AA; 36531 MW; 38E595564EDBF3D CRC64;

Query Match 11.0%; Score 180; DB 2; Length 327;
Best local Similarity 23.8%; Pred. No. 9.3e-08;
Matches 76; Conservative 56; Mismatches 146; Indels 42; Gaps 13;

QY 5 VPTFSLAELOQGL--HODEF---RRCLRDKGLFYLTDCGLDTELSKADVIDPFEH 58
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 18 VPTLSIASYTKGTAGRSSEFIGKLFYGLKEYGFIIKDHNVQAAADLHKAVIDIKNFYALP 77

QY 59 SEAEKRAVTSVPVTMRGFT---GLSESTAOITNTGSYSDYSGVSMGT-----ADNLF 110
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 78 TEV-KSYISIPKAGFQGTTPFGQEHAKDSPVMDLKEFMHVGRELSGNPLKAVIPANW 136

QY 111 PS--GDFERIWTQYFDRQYTSRAVAREVLRA---TGTERPDGVAEFLDCEPLAFRYF 164
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 137 PSELPEFQ---SHFSKLVAALEAGDVMLLELTMPLVEKDFPAKWTGDGNSIILAHY 192

QY 165 PQVPEHRSABEOPLRMAPHYDLSMTLLIOQTPCANCFVSLQAEVGAFDLP 224
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 193 PPIPE--GVDPRCVRAAHEDINFTIL--PAATSGQLKXDRDSTWLDIDSEPTLLIV 247

QY 225 FCGAIVTLVGGGVKAPRHVVAAPRRDQIAGSRTSSVFPLRNADFTS-VPLARECGF 283
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 248 DVDMDLARLTNDVLPSTHRAVINPDG--TNGSRYSMPFFMHPHPLAMLSCLP----- 298

QY 284 DVSIDGETATFDPMIGNTV 303
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 299 --SKGTGAKVADITGODFL 316

Search completed: September 19, 2005, 15:18:14
Job time : 96 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2005, 15:06:04 : Search time 24 Seconds
(without alignment)
967.327 Million cell updates/sec

Title: 10719236-1_232-1164

Perfect score: 1636

Sequence: 1 MDTVPTFSIAELQOGLHOD.....ATPDWIGNVNIRTSKA 311

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PCITS_COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1629	99.6	311	3	US-09-413-231-18
2	1210	74.0	313	1	US-08-592-411-15
3	1195.5	73.1	314	1	US-08-592-411-17
4	972.5	59.4	310	5	PCT-US95-04801-6
5	168	10.3	329	3	US-09-413-231-5
6	168	10.3	329	3	US-09-413-231-9
7	165	10.1	333	2	US-08-379-556A-10
8	153	9.4	333	2	US-08-379-556A-8
9	149.5	9.1	327	4	US-09-902-540-15162
10	147	9.0	333	4	US-09-413-231-10
11	143	8.7	347	2	US-08-379-556A-2
12	133.5	8.2	331	3	US-09-413-231-4
13	133.5	8.2	343	3	US-09-454-034-8
14	129	7.9	321	3	US-09-413-231-8
15	122.5	7.5	331	3	US-09-413-231-1
16	122.5	7.5	331	3	US-09-413-231-2
17	122.5	7.5	331	3	US-09-413-231-3
18	122.5	7.5	386	2	US-08-553-367A-2
19	122.5	7.5	386	4	US-09-295-306-2
20	122.5	7.5	386	4	US-09-734-719-2
21	122.5	7.5	386	4	US-09-371-307-78
22	121	7.4	378	2	US-08-553-367A-6
23	121	7.4	378	3	US-09-295-306-6
24	121	7.4	378	4	US-09-734-719-6
25	120.5	7.4	389	4	US-09-248-796A-14285
26	119	7.3	351	4	US-09-489-039A-9415
27	117	7.2	238	4	US-09-270-767-45029

28	116	7.1	344	4	US-09-252-991A-28503	Sequence 28503, A
29	112.5	6.9	377	2	US-08-553-367A-4	Sequence 4, Appli
30	112.5	6.9	377	4	US-09-295-306-4	Sequence 4, Appli
31	112.5	6.9	377	4	US-09-734-719-4	Sequence 20513, A
32	108.5	6.6	390	4	US-09-252-991A-20513	Sequence 7, Appli
33	108	6.6	338	3	US-09-413-231-7	Sequence 4, Appli
34	107	6.5	241	2	US-08-379-556A-4	Sequence 6, Appli
35	105	6.4	338	3	US-09-413-231-6	Sequence 59, Appli
36	104	6.4	321	4	US-09-371-307-59	Sequence 38, Appli
37	103.5	6.3	331	4	US-09-614-912-44	Sequence 40, Appli
38	103.5	6.3	368	4	US-09-614-912-38	Sequence 1631, A
39	103.5	6.3	734	4	US-09-248-796A-16776	Sequence 2, Appli
40	102.5	6.3	296	4	US-09-614-912-40	Sequence 17497, A
41	102.5	6.3	370	4	US-09-614-912-2	Sequence 10, Appli
42	101.5	6.2	403	4	US-09-248-796A-16531	
43	100	6.1	331	4	US-09-719-108-2	
44	96	5.9	453	4	US-09-248-796A-17497	
45	94.5	5.8	112	3	US-09-454-034-10	

ALIGNMENTS

```
RESULT 1
US-09-413-231-18
; Sequence 18, Application US/09413231
; Patent No. 6284483
; GENERAL INFORMATION:
; APPLICANT: Dilley, David R
; APPLICANT: Kadyrzhanova, Dina K
; APPLICANT: Wang, Zhenyong
; APPLICANT: Warner, Toni M
; TITLE OF INVENTION: Modified Synthetases To Produce Penicillins and
; TITLE OF INVENTION: Cephalosporins Under the Control of Bicarbonate
; FILE REFERENCE: MSU41-453
; CURRENT APPLICATION NUMBER: US/09/413,231
; CURRENT FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: modified DAOCS
; NAME/KEY: MUTAGEN
; LOCATION: (181)
; OTHER INFORMATION: Ala181 in native DAOCS modified to Arg
US-09-413-231-18

Query Match          99.6%; Score 1629; DB 3; Length 311;
Best Local Similarity 99.4%; Pred. No. 5e-174;
Matches 309; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDTVPTFSIAELQOGLHODFFRCLRDKGLFYLTDDGLDTEKSKNDVIIDFFEGSE 60
   |||||||
DB 1 MDTVPTFSIAELQOGLHODFFRCLRDKGLFYLTDDGLDTEKSKNDVIIDFFEGSE 60
   |||||||

QY 61 AEKAAVSPVPTMGRGTGLESESTAOITNTGSGSDVSMCSKMTAANLFPSGDFERTWT 120
   |||||||
DB 61 AEKAAVSPVPTMGRGTGLESESTAOITNTGSGSDVSMCSKMTAANLFPSGDFERTWT 120
   |||||||

QY 121 QYFPRQYTAASRAVAREVLRATGTETPDGVEAFIDCEPLLRFRYPQVPEHRSABEQPLRM 180
   |||||||
DB 121 QYFPRQYTAASRAVAREVLRATGTETPDGVEAFIDCEPLLRFRYPQVPEHRSABEQPLRM 180
   |||||||

QY 181 APHYDLSMVTLLIOOTPCANGFVSLQAEVGAFTDLPYRPAVLYFGAIAITLVYGGQVKA 240
   |||||||
DB 181 RPHYDLSMVTLLIOOTPCANGFVSLQAEVGAFTDLPYRPAVLYFGAIAITLVYGGQVKA 240
   |||||||

QY 241 PRHHVAAPRRDQIGSSRTSSVFPLRNADFTFSVPLARECGFVSLDGETATPQDDIGG 300
   |||||||
```

Db 241 PRHHVAAPRRDQIAGSSRTSSVFLRPVADFTFVPLARECGFVDSLDEGTATFQDWIGG 300
Qy 301 NYNIRRTSKA 311
Db 301 NYNIRRTSKA 311

RESULT 2

US-08-592-411-15
Sequence 15, Application US/08592411
Patent No. 5726032
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Process for the Efficient Production of
TITLE OF INVENTION: 7-ADCA via 2-(Carboxymethylthio)acetyl-7-ADCA and
TITLE OF INVENTION: 3-(Carboxymethylthio)propionyl-7-ADCA
NUMBER OF SEQUENCES: 17
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,411
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 313 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-592-411-15

Query Match 74.0%, Score 1210, DB 1, Length 313;
Best Local Similarity 71.0%; Pred. No. 5,1e-127;

Matches 220; Conservative 40; Mismatches 50; Indels 0; Gaps 0;

Qy 2 DTTVPPTSLAELOGGHODEFRCLDKGLFYITDCGLTDTELKSAKDVIDFEHSGSEA 61
Db 3 DATVPFDLAELREGHDEFRCLREKGVLYLKGTLAEADHSAEIAVDFPDHGTGEA 62
Qy 62 EKRAVTSVPVPMRGRFTGLSESTAOITNTGSDYSNCSYMGTDNLFPSSGDFERITWTQ 121
Db 63 EKXAVMTPIPTIRRGVAGLESESTAOITNTGKYTDYSMSYMGTDNLFPSEAEKXAMED 122
Qy 122 YFDRQYTSRAVAREVLATGTEPDGVEAFLDCEPLLRFYFPQVPEHRSAGEOPLRMA 181
Db 123 YFARMYRASQVAVARQVLTSGAEPEVGMDFLDCEPLLRLRYFPVEVEDRVAEEQPLRMA 182
Qy 182 PHYLSMWTLLIOOTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVTTGGQVKAP 241
Db 183 PHYLSITVTLHQTPCANGFVSLQAEVGVSDVDIPAGQAVLVFCGAAVTLVADGAIKAP 242
Qy 242 RHVVAAPRRDQIAGSSRTSSVFLRPVADFTFVPLARECGFVDSLDEGTATFQDWIGG 301
Db 243 KHHVAAAPGADKRVGSSRTSSVFLRPVADFTFVPLARECGFVDSLDEGTATFQDWIGG 302
Qy 302 YNIRRTSKA 311
Db 303 YINIRKTA 312

RESULT 3

US-08-592-411-17
Sequence 17, Application US/08592411
Patent No. 5726032
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Process for the Efficient Production of
TITLE OF INVENTION: 7-ADCA via 2-(Carboxymethylthio)acetyl-7-ADCA and
TITLE OF INVENTION: 3-(Carboxymethylthio)propionyl-7-ADCA
NUMBER OF SEQUENCES: 17

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,411
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 314 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-592-411-17

Query Match 73.1%, Score 1195.5, DB 1, Length 314;
Best Local Similarity 70.4%; Pred. No. 2,2e-125;

Matches 219; Conservative 40; Mismatches 51; Indels 1; Gaps 1;

Qy 2 DTTVPPTSLAELOGGHODEFRCLDKGLFYITDCGLTDTELKSAKDVIDFEHSGSEA 60
Db 3 DATVPFDLAELREGHDEFRCLREKGVLYLKGTLAEADHSAEIAVDFPDHGTGE 62
Qy 61 AEKRAVTSVPVPMRGRFTGLSESTAOITNTGSDYSNCSYMGTDNLFPSSGDFERITWT 120
Db 63 AEKXAVMTPIPTIRRGVAGLESESTAOITNTGKYTDYSMSYMGTDNLFPSEAEKXAME 122
Qy 121 QYFDRQYTSRAVAREVLATGTEPDGVEAFLDCEPLLRFYFPQVPEHRSAGEOPLRM 180
Db 123 DYFARMYRASQVAVARQVLTSGAEPEVGMDFLDCEPLLRLRYFPVEVEDRVAEEQPLRM 182
Qy 181 APHYLSMWTLLIOOTPCANGFVSLQAEVGAFTDLPYRPDAVLVFCGAIATLVTTGGQVKAP 240
Db 183 APHYLSITVTLHQTPCANGFVSLQAEVGVSDVDIPAGQAVLVFCGAAVTLVADGAIKAP 242
Qy 241 PRHHVAAPRRDQIAGSSRTSSVFLRPVADFTFVPLARECGFVDSLDEGTATFQDWIGG 300
Db 243 KHHVAAAPGADKRVGSSRTSSVFLRPVADFTFVPLARECGFVDSLDEGTATFQDWIGG 302
Qy 301 NYNIRRTSKA 311
Db 303 NYINIRKTA 313

RESULT 4

PCT-US95-04801-6
Sequence 6, Application PC/TUS9504801
GENERAL INFORMATION:
APPLICANT: Martin, Juan F.
APPLICANT: Coque, Juan R.
APPLICANT: Enguita, Francisco J.
APPLICANT: Fuente, Juan L.
APPLICANT: Liras, Paloma
TITLE OF INVENTION: DNA ENCODING CEPHAMYCIN BIOSYNTHESIS
TITLE OF INVENTION: LATE GENES
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: John W. Wallen III
STREET: P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04801

FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Wallen III, John W.
REGISTRATION NUMBER: 35,403
REFERENCE/DOCKET NUMBER: 19179
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3905
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-04801-6

Query Match 59.4%; Score 972.5; DB 5; Length 310;
Best Local Similarity 61.5%; Pred. No. 2,2e-100;
Matches 187; Conservative 37; Mismatches 79; Indels 1; Gaps 1;

QY 2 DTTVPFSLAELOQGLHODEFRRCILDKGLFYLTDCGLTDTLKSAXDIVIDFFEHGSEA 61
DB 3 DKTVPFVSMALRDGSRDEFRMAR-RGVFYLTGATERDHRVATDTAMDFFAQGTAE 61
QY 62 EKRATSPVPTMRGRFTGLESESTAOITNTGYSYDYSKMGITADNLPFGSDFERTITQ 121
DB 62 EKQAVTTRKVTMRKGSALAEASTAOVNTGTYTYSMSYSGIGNLPFSKFFSSWTD 121
QY 122 YEDROYTASRAVAREVLRATGTEPDGVEAFDCEPLRFYRFPVPEHRSABEQPLMA 181
DB 122 YFDSLXYRAAQERARLVLTAGTYDGEDDLTDLCBPVLRKRFPEVPEHRAEYEPKMA 181
QY 182 PHYDSLMTLLIQOTPCANGFVSLQAEVGAFTDLFYPKPAVLVFCGAIATVITGQVAP 241
DB 182 PHYDSLITFIHQTPCANGFVSLQAEVGDENVSLPHVEAVVVLGCAIAPLVTCGAVPAP 241
QY 242 RHVVAAPRRDQIAGSSRTSSVFELPRNADFTTSVPLARBCGDSVLDGTAFTQWIGN 301
DB 242 NHHVVSPPASMLKSGDRITSSVFELRPSDTFTSPVDARKYGLVDLMEKATFGMIGTN 301
QY 302 YVNI 305
DB 302 YVTM 305

RESULT 5
US-09-413-231-5
Sequence 5, Application US/09413231

GENERAL INFORMATION:
PATENT NO. 6284483
APPLICANT: Dilley, David R
APPLICANT: Kadyrzhanova, Dina K
APPLICANT: Wang, Zhenyong
APPLICANT: Warner, Toni M
TITLE OF INVENTION: Modified Synthetases To Produce Penicillins and
TITLE OF INVENTION: Cephalosporins Under the Control of Bicarbonate
FILE REFERENCE: MSU41-453
CURRENT APPLICATION NUMBER: US/09/413,231
CURRENT FILING DATE: 1999-10-06
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 5
LENGTH: 329
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: modified IPNS
OTHER INFORMATION: from Streptomyces clavuligerus
NAME/KEY: MUTAGEN
LOCATION: (210)
OTHER INFORMATION: Glu210 in native IPNS modified to Arg

US-09-413-231-5

Query Match 10.3%; Score 168; DB 3; Length 329;
Best Local Similarity 23.8%; Pred. No. 4.2e-10;
Matches 77; Conservative 50; Mismatches 150; Indels 46; Gaps 14;

QY 5 VPTFSLAE-----OQGLHODEFRRCILDKGLFYLTDCGLTDTLKSAXDIVIDFFEHG 58
DB 10 VPTIDISPLFTDAAAKRAVAEIHGACRGSGFFYATNHGV---DVQQLDVNVEFHGM 66
QY 59 SEAEK-----RAVTSVPVPMRGRF---TGLESESTAOITNTGYSYDYSKMGIT---A 106
DB 67 TDQEKHDLAIHAVPNDPHVANGYKAVPGKAVESFCYLPNDPGEDEHPM-IAAGTPME 125
QY 107 DNLFPSSG---FERITQYDROYTASRAVAREVLRATGTEPDGVEA-----FLDCE 156
DB 126 VNLMPDEERHPRFPFCGKYRQMLKSTVMRGLALAG-RPHFPDAALAEODSLSSV 184
QY 157 PLRFRYPPOVPEHRSABE-OPLRMAPHYDSLMTLLIQOTPCANGFVSLQAEVGAFTDL 215
DB 185 SLIRPYLAEYPPVKTGPDGLSFRDHDVSMITVLFTQYQV---LQVETDGRDI 240
QY 216 PYRPDAVLVFCGAIATVITGQVAPRRHVAAPRRDQIAGSSRTSSVFELPRNADFTTSV 275
DB 241 FTSENDVLNCGTYMAVHTNDYFPAPNHRV-----KFNARLRLPFLNGCHBAVTE- 293
QY 276 PLARECGPDSVLDGFTATFQDWI 298
DB 294 PFVPE-GASEVRNEALSYGDYL 315

RESULT 6
US-09-413-231-9
Sequence 9, Application US/09413231

GENERAL INFORMATION:
PATENT NO. 6284483
APPLICANT: Dilley, David R
APPLICANT: Kadyrzhanova, Dina K
APPLICANT: Wang, Zhenyong
APPLICANT: Warner, Toni M
TITLE OF INVENTION: Modified Synthetases To Produce Penicillins and
TITLE OF INVENTION: Cephalosporins Under the Control of Bicarbonate
FILE REFERENCE: MSU41-453
CURRENT APPLICATION NUMBER: US/09/413,231
CURRENT FILING DATE: 1999-10-06
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 9
LENGTH: 329
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: modified IPNS
OTHER INFORMATION: from Streptomyces clavuligerus strain NRRL 3585
NAME/KEY: MUTAGEN
LOCATION: (210)
OTHER INFORMATION: Glu210 in native IPNS modified to Arg

Query Match 10.3%; Score 168; DB 3; Length 329;
Best Local Similarity 23.8%; Pred. No. 4.2e-10;
Matches 77; Conservative 50; Mismatches 150; Indels 46; Gaps 14;

QY 5 VPTFSLAE-----OQGLHODEFRRCILDKGLFYLTDCGLTDTLKSAXDIVIDFFEHG 58
DB 10 VPTIDISPLFTDAAAKRAVAEIHGACRGSGFFYATNHGV---DVQQLDVNVEFHGM 66
QY 59 SEAEK-----RAVTSVPVPMRGRF---TGLESESTAOITNTGYSYDYSKMGIT---A 106
DB 67 TDQEKHDLAIHAVPNDPHVANGYKAVPGKAVESFCYLPNDPGEDEHPM-IAAGTPME 125
QY 107 DNLFPSSG---FERITQYDROYTASRAVAREVLRATGTEPDGVEA-----FLDCE 156

Db 126 UNLMDBERHRFRPFCEGYRQMLKSLVLMRGALALG-RPEHFDALAEODSLSSV 184
Qy 157 PLRFRFPPOVPEHRSAE-OPLMAMPYDLSMTLLIQTPCANGFVSLQAEVGCATDL 215
Db 185 SLIRYPLIEEYPKCTGPDGLSFRHLDVDSMITVLFQTOVON----LQVETDGMRI 240
Qy 216 PYRDAVLVFCGALATLVGGQVAPRHHVAAPRRDQIAGSSRFSVFPLRPNADFTFSV 275
Db 241 PTSNDPLVNCGTMAHTNDYPPAPNHRV-----KFAVAERLSLFFPLANGHEAVIE- 293
Qy 276 PLARECGFVSLDEGTATFODMI 298
Db 294 PFVPE-GASEVRNEALSVDYL 315

RESULT 7
US-08-379-556A-10
; Sequence 10, Application US/08379556A
; Patent No. 5859329
; GENERAL INFORMATION:
; APPLICANT: HOLTON, TIMOTHY A.
; APPLICANT: KEAM, LISA A.
; TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOL
; TITLE OF INVENTION: SYNTHASE ENZYMES AND USES THEREFORE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
; STREET: 400 GARDEN CITY PLAZA
; CITY: GARDEN CITY
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/379,556A
; FILING DATE: 22-MAR-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: DIGILIO, FRANK S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 9592
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516)742-4343
; TELEFAX: (516)742-4366
; TELEEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 335 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-379-556A-10

Query Match 10.1%; Score 165; DB 2; Length 335;
Best Local Similarity 22.8%; Pred. No. 9.3e-10;
Matches 72; Conservative 55; Mismatches 135; Indels 54; Gaps 14;
Qy 5 VPTFSIAELQOGLHODEFRRCCLRDKGLFYLTDCGLTDTLBSAKADIYIDPFHESBAKR 64
Db 43 VPVIDLSQADNESLVALISKASKDWGIFQVNNHGISPLISKLQNVGKEFFELFQF-EKE 101
Qy 65 AVTSPVPTMRGRGTGLESESTQAITNTGSSYSDYSWCMYSGTADNLF-----PSGDFRIW 119
Db 102 VIAKP-----DGYGVE-----GYCTKLQKEVQGGKGVNDLPHLWMPSPSTINQFW 148
Qy 120 TQ-----YFD-RQYTAAR-AVAREVL-----RATGTEPD-----GVBAFLDCPPLRFR 162
Db 149 PKNPPSRVDTWTEYTSILIGVANKLLGLSLGLEDEDEVKQALGGBDLIY---MLKIN 204

Qy 163 YEPQVEHRSABEQPLRMAPHYDLSMTVLLIQTPCANGFVSLQAEVGCATDLPYRDAV 222
Db 205 YIPPCP-----CEPLALGVAPHDTMSSITLL-----VNEVQGLQVFDGGMVDAVYIPNML 256
Qy 223 LVFCGAIATLVTTGGQVAPRHHVAAPRRDQIAGSSRFSVFPLRPNADFTFSVPLARECG 282
Db 257 IHHGQIELISNGKYSVYHRSTYVKE-----KTRMSWPAFLPPEFEVG-FIRKLVN 310
Qy 283 FVYSLDEGTATFODMI 298
Db 311 KDDPPKXKTKKXYDY 326

RESULT 8
US-08-379-556A-8
; Sequence 8, Application US/08379556A
; Patent No. 5859329
; GENERAL INFORMATION:
; APPLICANT: HOLTON, TIMOTHY A.
; APPLICANT: KEAM, LISA A.
; TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOL
; TITLE OF INVENTION: SYNTHASE ENZYMES AND USES THEREFORE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
; STREET: 400 GARDEN CITY PLAZA
; CITY: GARDEN CITY
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/379,556A
; FILING DATE: 22-MAR-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: DIGILIO, FRANK S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 9592
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516)742-4343
; TELEFAX: (516)742-4366
; TELEEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-379-556A-8

Query Match 9.4%; Score 153; DB 2; Length 333;
Best Local Similarity 21.6%; Pred. No. 2e-08;
Matches 71; Conservative 57; Mismatches 126; Indels 74; Gaps 15;
Qy 1 MDTT--VPTFSIAELQOGLHODEFRRCCL-----RDKGLFYLTDCGLTDTLBSAKADIYIDF 54
Db 39 LDTVLVEPALDLS-----LEEDDVVKLVLASAKEMGLFQVTNNGIPIETVYEKLQKGVKMF 93
Qy 55 FEHGSBAEKRAVTSPPVPTMRGRGTGLESESTQAITNTGSSYSDYSWCMYSGTADNLFPSGD 114
Db 94 FRAPAE-EKETLAKKPGVGVGTMLQKEIQGR-----KGVWDHL----- 132
Qy 115 FERITWT-----QYF-----DRQYASRAVAREVL-----RATGTEPDG-----GV 149
Db 133 FHKVPPSVVNTYVWVWTKTSYRANAEYTKYLRIYADKLFKCKSKGLGLEBDEVKSCGN 192
Qy 150 EAFLLDCPPLRFRYFPQVPEHRSABEQPLRMAPHYDLSMTVLLIQTPCANGFVSLQAEVGC 209

```
Db      193 E----DIVVLLKINYYPCPR-----PDIALGVAHTDLVITL-----VENDVAGLQVSRD 241
Qy      210 GAFDLPYRPDAVVFCCGAIATLVGGGVKAPRHVAAFRQIAGSSSTSVFLRPNA 269
Db      242 GFWYDKYIPNALIITHVGQMEIMSGEYKVALHSTVKE-----RTIRISMPVPLEPPS 296
Qy      270 DFTFSVPLARECGFVSLDGETATFQDM 297
Db      297 DFAVG-PIPKLISDEKPAKTKIVSEY 323

RESULT 9
US-09-540-15162
; Sequence 15162, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 15162
; LENGTH: 327
; TYPE: PRF
; ORGANISM: Myxococcus xanthus
US-09-902-540-15162

Query Match      9.1%; Score 149.5; DB 4; Length 327;
Best Local Similarity 21.6%; Pred. No. 4.9e-08;
Matches 77; Conservative 39; Mismatches 118; Indels 123; Gaps 15;

Qy      21 EERRCLRDKGLFYLTDCGLTDTL-----KSAADIY-----DFFEH 57
Db      19 EIEQCRSGSFYVGHGSGEVLARLERESHFPALPVAKAIALMSGVAMRGWPL 78
Qy      58 GSEAEKRAVTSVPFTRRG-FTGLESESTAOITNTG-----SYSDY- 97
Db      79 GGE-----LTSGRPARKSLYLTGLGSEHPRVXKGMPLHGANLMPAEVPELRAVLDYV 133
Qy      98 -----SVCYSMGCTADNLPFGSDFERIMTOYDROYTASRAVAREVLRATGEP 145
Db      134 AACTRAAHALMEGMALSLG-----LDADYFRRHHTADPTV----- 168
Qy      146 DGGVAFPLDCEPLRFRRYPQVBEHRSABEQPLRAAPHYDLSMTLLIOQTPCANGFVSLQ 205
Db      169 -----LRFIFHPAEPOH---EVSWGGEHTDGLTLAODD--NG--GLQ 209
Qy      206 AEVGGAFTDLPYRPDAVVFCCGAIATLVGGGVKAPRHVAAFRQIAGSSSTSVFL 265
Db      210 VTPRGWVEVPLPGTLVCNIGMDLDMTCGMYRSTPHRV-----KNVSGKDRLSFPLPF 264
Qy      266 RPNADFTTSV-PLARECGFV-----SLDGETATFQDMIGNYVNIIRTSK 310
Db      265 DP--DFAABVHPLPRGAGADVDDDRARRWDGASVHAFQYGVGYLLGVFKVLSRSR 319

RESULT 10
US-09-413-231-10
; Sequence 10, Application US/09413231
; Patent No. 6284483
; GENERAL INFORMATION:
; APPLICANT: Dille, David R
; APPLICANT: Kadyrzhanova, Dina K
; APPLICANT: Wang, Zhenyong
; APPLICANT: Warner, Toni M
; TITLE OF INVENTION: Modified Synthetases To Produce Penicillins and
; TITLE OF INVENTION: Cephalosporins Under the Control of Bicarbonate
```

```
; FILE REFERENCE: MSU41-453
; CURRENT APPLICATION NUMBER: US/09/413,231
; CURRENT FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10
; LENGTH: 333
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: modified IPNS
; NAME/KEY: MUTAGEN
; LOCATION: (214)
; OTHER INFORMATION: Glu214 in native IPNS modified to Arg
US-09-413-231-10

Query Match      9.0%; Score 147; DB 3; Length 333;
Best Local Similarity 23.2%; Pred. No. 9.6e-08;
Matches 70; Conservative 37; Mismatches 131; Indels 64; Gaps 13;

Qy      5 VPTSLAELOQGLHOD-----EPRCLRDKGLFYLTDCGLTDTLKSADIYIDFEH 57
Db      10 VPTIDISPL-FGTPDPAKAVARQINEACRSGSGFYVASHHG1--DVRLQDVNVEFRT 65
Qy      58 GSEAEKRAVTSVPFTRRG-FTGLESESTAOITNTGSDYSDYSCY---SMGTADNL 109
Db      66 MTQDEKIDLAITHAVNENNSHVRNGY-----YMARPGKTVESWCYLNPSFGEDHPM 116
Qy      110 FPGS-----DPERIMTOYDROYTASRA--VAREVLBATGT-----EPDG 147
Db      117 IKAGTPMHEVNVWPDERRHPDRFSGQYREVRRLSKVLLRFPALALGPEEFENEV 176
Qy      148 GVEAFDCEPL-LRFRRYPQVBE--HRSABEQPLRAAPHYDLSMTLLIOQTPCANGFVS 203
Db      177 TEEDTLSCRSIMIRYPYLDYPEBAIKTGPGRSLSPRDHLDVSMITVLFTQTEVON---- 232
Qy      204 LQAEVGAFTDLPYRPDAVVFCCGAIATLVGGGVKAPRHVAAFRQIAGSSSTSVF 263
Db      233 LQVETVDGWSLPTSGENFLNCGTYLGYLTNDYFPAPNRHV-----KYNAERLSLPF 286
Qy      264 FL 265
Db      287 FL 288

RESULT 11
US-08-379-556A-2
; Sequence 2, Application US/08379556A
; Patent No. 5859329
; GENERAL INFORMATION:
; APPLICANT: HOLTON, TIMOTHY A.
; APPLICANT: KEAM, LISA A.
; TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOL
; TITLE OF INVENTION: SYNTHASE ENZYMES AND USES THEREFORE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
; STREET: 400 GARDEN CITY PLAZA
; CITY: GARDEN CITY
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/379,556A
; FILING DATE: 22-MAR-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
```

NAME: DIGIGLIO, FRANK S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 9592
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516)742-4343
TELEFAX: (516)742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 347 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-379-556A-2

Query Match 8.7%; Score 143; DB 2; Length 347;
Best Local Similarity 22.3%; Pred. No. 2.9e-07;
Matches 59; Conservative 43; Mismatches 14; Indels 48; Gaps 9;

QY 27 RDKGLFYLTDCGLTDTLTKSAKDVIIDPFHSGSEAEKRAVT-SPVPTMRGF-TGLESES 84
DB 77 KEMGIFQILNHGIDPEADIAIDQKXKFFEHVPEKEKELIAKTPGSNDIBGYSLOKEV 136
QY 85 TAOITNTGYSYSDYSMCSMTADNLF-----PSGDFERIW-----TOYFDRQYT 128
DB 137 EKG-----KGMVDHLFKHIMPSPSAVNYRYPKNDPSYREANEEYGGKMR 181
QY 129 ASRAVAREVLARATGTEPDCGVEALDGE--PLAFRFRFPQVPEKRSABEOPLRAPVYDL 186
DB 182 VDRIFFKSLSLGIGLEGHMELEAGDEIVYLKINYPGPR---PDIALGVVAHTDM 237
QY 187 SMVLLIOOTPCANGFVSLQAEVGAFTDLPYRPDAVIFGCAIATLVYGGVKAPRHVA 246
DB 238 SYITIL-----VPNEVQGLQVFKDGMWDVKITPAILVHIGDVEIISNGKYSVYHRTT 293
QY 247 APRRDQIAGSSRTSSVFFLRPNAD 270
DB 294 VNK-----DKTRMSWPVFLPEPSE 312

RESULT 12
US-09-413-231-4
; Sequence 4, Application US/09413231
; Patent No. 6284483
; GENERAL INFORMATION:
; APPLICANT: Dille, David R
; APPLICANT: Kadyrzhanova, Dina K
; APPLICANT: Wang, Zhenyong
; APPLICANT: Warner, Toni M
; TITLE OF INVENTION: Modified Synthetases To Produce Penicillins and
; FILE REFERENCE: MEU41-453
; CURRENT APPLICATION NUMBER: US/09/413,231
; CURRENT FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: modified IPNS
; OTHER INFORMATION: from Penicillium chrysogenum
; NAME/KEY: MUTAGEN
; LOCATION: (212)
; OTHER INFORMATION: Glu212 in native IPNS modified to Arg
US-09-413-231-4

Query Match 8.2%; Score 133.5; DB 3; Length 331;
Best Local Similarity 23.8%; Pred. No. 3.1e-06;
Matches 76; Conservative 40; Mismatches 13; Indels 73; Gaps 14;
QY 27 RDKGLFYLTDCG-----LTDTELKSAKDVIIDPF--EHGSE----- 60

DB 37 RDGFFYAVNHGADVRLSKTRFEHFSITDE---KMLAIRANKKHQDDQIRAGYIS 93
QY 61 -AEKRAVTSBPVPMRGRFTGLESESTAOITNTGYSYSDYSMCSMTAD--NLFP-----G 113
DB 94 IPEKAVES-----FCYL-----NPFKDHPLIQSKTPTHEANVWPDEKKHP 136
QY 114 DFERIMTOYFDRQYTSRAVAREVLARATGTEPDCGVEAF-----LDCEPLLRYPFQV 168
DB 137 GFREPAEQIYWDVFGSSALRGYALALGKEEDFESRHFKEDALSSVYLIRIPLYNPIL 196
QY 169 EH--RSABE-QPLRMAPHYDLSMTLLIOOTPCANGFVSLQAEVGAFTDLPYRPDAVLYE 225
DB 197 PAIKTADGDTKLSFMWHEVDLSLITVLYOSDVAN---LQVEWPGQYLDIADNADNLYVN 252
QY 226 CGAIALVYGGVKAARRHVAAPRRDQIAGSSRTSSVFFLRPNADTFESVPLAREGFPV 285
DB 253 CGSYMAHITNNYYPADIRV-----KWNBERQSLPFFV--NLGFNDTVQWPDPSEKDG 304
QY 286 SLDEGTATPDMDIGAGNYVNI 305
DB 305 KTDQRPISGDIQLNGVLSL 324

RESULT 13
US-09-454-034-8
; Sequence 8, Application US/09454034
; Patent No. 6380464
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Fader, Gary M.
; TITLE OF INVENTION: Plant Flavonol Synthase Homologs
; FILE REFERENCE: BB1298 US NA
; CURRENT APPLICATION NUMBER: US/09/454,034
; CURRENT FILING DATE: 1999-12-03
; EARLIER APPLICATION NUMBER: 60/110,875
; EARLIER FILING DATE: 1998-December-04
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Glycine max
US-09-454-034-8

Query Match 8.2%; Score 133.5; DB 3; Length 343;
Best Local Similarity 22.7%; Pred. No. 3.3e-06;
Matches 65; Conservative 46; Mismatches 126; Indels 49; Gaps 13;

QY 5 VPTFSAELQOQGHODEFRRCLRDKGLFYLTDCGLTDTLTKSAKDVIIDPFHSGSEAEKR 64
DB 51 VPIIDSDPDEGVNHEILEASRDQMFQIVNHDISDVIRKLGQSGKFFELPQE-EKE 109
QY 65 AVTSVPVPMRGRFTGLESESTAOITNTGYSYSDYSMCSMTADNLF-----PSGDFERIW 119
DB 110 LIAKP-----AGSDS-----IEGYTKLQKEVNGKGMVDHLFIHWPSSINYSFW 156
QY 120 TO-----YEDROYTS-BAVAEVLRA-----TGTE-----PQCGVEAPLDCBPLARFY 164
DB 157 PQNPSTREYNEBYCNHKGVDKLFKSMVSGLEENELKSGAND--DMHYLLKINY 214
QY 165 PQVPEHRSABEOPLRMAPHYDLSMTLLIOOTPCANGFVSLQAEVGAFTDLPYRPDAVLY 224
DB 215 PRCP-----CPDLVLSGPRPHDMSYLLIL-----VPNEVQGLQACRDBHWDVKVPAALVI 266
QY 225 FCGAIALVYGGVKAARRHVAAPRRDQIAGSSRTSSVFFLRPNAD 270
DB 267 HIGDQWEIISNGKXKAVFHRITV-NKDE-----TRMSWPVFIPEKKE 307

RESULT 14
US-09-413-231-8
; Sequence 8, Application US/09413231

```
/ Patent No. 6284483
/ GENERAL INFORMATION:
/ APPLICANT: Dilley, David R
/ APPLICANT: Kadyrtzanova, Dina K
/ APPLICANT: Wang, Zhenyong
/ APPLICANT: Warner, Toni M
/ TITLE OF INVENTION: Modified Synthetases To Produce Penicillins and
/ TITLE OF INVENTION: Cephalosporins Under the Control of Bicarbonate
/ FILE REFERENCE: MSU41-453
/ CURRENT APPLICATION NUMBER: US/09/413,231
/ CURRENT FILING DATE: 1999-10-06
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 8
/ LENGTH: 321
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: modified IPNS
/ NAME/KEY: MUTAGEN
/ LOCATION: (211)
/ OTHER INFORMATION: Glu211 in native IPNS modified to Arg
US-09-413-231-8

Query Match
Best Local Similarity 7.9%; Score 129; DB 3; Length 321;
Matches 72; Conservative 32; Mismatches 150; Indels 62; Gaps 12;

QY 5 VPTFSIAELIQGLHODEFR-----RCLRDKGLFYLTDCGLTDELKSAKDVIDFFEH 57
DB 10 VPTLISPOLFGTDPPRTSRGRSRPARSGGFPAASHHGIDVRLQWNSN----- 61
QY 58 GSEAEKRAVTSVPTMRRCFTGLESESTAQ---ITNTGSYSDYSNCT--SMGTADNLF 110
DB 62 ----ESTWTQDRSTWRSTRYNNENSHVNGYUAPRERETVESMCYLNPSFGEDHPMM 117
QY 111 PSG-----DPERIMTQYDRQYTASRAV---AREVLRAATGNEPRGGV-- 149
DB 118 KAGTGMHEVNVWPDERRHDPFGSGEQY-HREVSASRCCCGASRRRROAGSSSNEVTE 176
QY 150 EAFLLDCEPLLRPRYPPOVE---HRSABEQPLRMAPHYDLSWVTLIQTPCANGFVSLQA 206
DB 177 EDTLSAVSMIRPYLDYPEALIKTGPDOGTRLSFRDHLDSMITVLSTKEVQN----LQY 232
QY 207 EVGGAFTDLPRPDALVFCGAIATLVGQYKAPRHVVAAPRDQIAGSSRTSSVFFLR 266
DB 233 ETVDDGQSLPTSGENFLINCGTYLITNDYFPAPNHRV-----KYVNAERLSLPFLH 286
QY 267 PNADFTFSVPLARECG 282
DB 287 AGQNSVMK-PFTRRTG 301

RESULT 15
US-09-413-231-1
/ Sequence 1, Application US/09413231
/ Patent No. 6284483
/ GENERAL INFORMATION:
/ APPLICANT: Dilley, David R
/ APPLICANT: Kadyrtzanova, Dina K
/ APPLICANT: Wang, Zhenyong
/ APPLICANT: Warner, Toni M
/ TITLE OF INVENTION: Modified Synthetases To Produce Penicillins and
/ TITLE OF INVENTION: Cephalosporins Under the Control of Bicarbonate
/ FILE REFERENCE: MSU41-453
/ CURRENT APPLICATION NUMBER: US/09/413,231
/ CURRENT FILING DATE: 1999-10-06
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 1
/ LENGTH: 331
/ TYPE: PRT
```

```
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: modified IPNS
/ OTHER INFORMATION: from Emericella nidulans strain D1A1
/ NAME/KEY: MUTAGEN
/ LOCATION: (212)
/ OTHER INFORMATION: Glu212 in native IPNS modified to Arg
US-09-413-231-1

Query Match
Best Local Similarity 7.5%; Score 122.5; DB 3; Length 331;
Matches 68; Conservative 45; Mismatches 129; Indels 83; Gaps 13;

QY 27 RDKGLFYLTDCGLTDELKSAK-----DIVIDPF--EHGSEAEKRAVTSVVP 71
DB 37 RDTGFFAVNHGINVQRLSQKTKEPHNSITPEEKMDLAIKAYNKEHODQ----- 85
QY 72 TMRGFTGLESESTAQITNTGSYSDYSNCT--SMGTAD-----NLPPS----- 112
DB 86 -VRAGY-----YLSIPGKAVESFCYLNPNFTPDHRIQAKTPTHEVNVWPDERTKH 135
QY 113 GDFERIMTQYDRQYTASRAVAREVLRAATGE-----PDGVEAFLDCEPLLRPRY 163
DB 136 PGFODFAEQYWDVDFGLSSALLKGYALALGKENPFARHFRKDDTLASVV---LIRPY 191
QY 164 FPQVPE---HRSABEQPLRMAPHYDLSWVTLIQTPCANGFVSLQAEVGAFTDLPRPD 220
DB 192 LDYPEAAIKTAADGQTKSFRMHEDVSLITVLQYSNVN---LQVETAGYQIDLEADT 247
QY 221 AVLFCGAIATLVGQYKAPRHVVAAPRDQIAGSSRTSSVFFLRPNADFTFSVPLARE 280
DB 248 GYLINGSYMALHTNNYKAPIHRV-----KWNNAERQSLPFVNLGYDSVIDPFDRE 301
QY 281 CGFDVSLDGETATFQDWIGNVTYNI 305
DB 302 P--NGKSDREPLSYGDYLLQNGLVS 324
```

Search completed: September 19, 2005, 15:19:20
Job time : 25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2005, 15:16:46 ; Search time 85 Seconds
(without alignments)
1481.559 Million cell updates/sec

Title: 10719236-1_232-1164

Perfect score: 1636
Sequence: 1 MDTVPTFTSLAEIQGLHQD.....ATFDWIGMGNVYINRTSKA 311

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1812044 seqs, 404927589 residues

Total number of hits satisfying chosen parameters: 1812044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
19: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1629	99.6	311	US-09-924-841-18	Sequence 18, Appl
2	173.5	10.6	345	US-10-732-923-9965	Sequence 9965, Ap
3	171	10.5	342	US-10-732-923-9978	Sequence 9978, Ap
4	169	10.3	366	US-10-369-493-12239	Sequence 12239, A
5	169	10.3	366	US-10-732-923-10056	Sequence 10056, A
6	168	10.3	329	US-09-924-841-5	Sequence 5, Appl
7	168	10.3	329	US-09-924-841-9	Sequence 9, Appl
8	168	10.3	412	US-10-732-923-9964	Sequence 9964, Ap
9	156	9.5	342	US-10-425-115-22033	Sequence 22033, A
10	156	9.5	375	US-10-425-114-58493	Sequence 58493, A
11	155	9.5	352	US-10-425-114-63175	Sequence 63175, A

12	154.5	9.4	366	US-10-437-963-146275	Sequence 146275,
13	154	9.4	342	US-10-425-115-22032	Sequence 22032,
14	154	9.4	371	US-10-425-114-62424	Sequence 62424, A
15	154	9.4	372	US-10-425-114-53454	Sequence 53454, A
16	154	9.4	372	US-10-425-114-72059	Sequence 72059, A
17	153	9.4	350	US-10-437-963-136698	Sequence 136698,
18	148.5	9.1	327	US-10-369-493-14043	Sequence 14043, A
19	148.5	9.1	344	US-10-425-115-316391	Sequence 316391,
20	148.5	9.1	356	US-10-424-599-258352	Sequence 258352,
21	147	9.0	333	US-09-924-841-10	Sequence 10, Appl
22	146.5	9.0	348	US-10-437-963-143438	Sequence 143438,
23	145	8.9	362	US-10-437-963-181202	Sequence 181202,
24	144.5	8.8	382	US-10-437-963-148589	Sequence 148589,
25	144.5	8.8	389	US-10-481-381-4	Sequence 4, Appl
26	143.5	8.8	342	US-10-767-701-44976	Sequence 44976, A
27	143	8.7	359	US-10-136-444-8	Sequence 8, Appl
28	140.5	8.6	428	US-10-437-963-176920	Sequence 176920,
29	139.5	8.5	380	US-10-431-273-88	Sequence 88, Appl
30	139.5	8.5	386	US-10-425-115-347325	Sequence 347325,
31	136	8.3	611	US-10-437-963-104090	Sequence 104090,
32	135.5	8.3	259	US-10-369-493-17939	Sequence 17939, A
33	134.5	8.2	405	US-10-739-930-77373	Sequence 77373, Ap
34	134	8.2	341	US-10-437-963-140887	Sequence 140887,
35	133.5	8.2	331	US-09-924-841-4	Sequence 4, Appl
36	133.5	8.2	334	US-10-424-599-190993	Sequence 190993,
37	133.5	8.1	350	US-10-425-115-254679	Sequence 254679,
38	131.5	8.0	339	US-10-369-493-7649	Sequence 7649, Ap
39	131	8.0	334	US-10-732-923-10146	Sequence 10146, A
40	131	8.0	356	US-10-425-114-64811	Sequence 64811, A
41	131	8.0	356	US-10-739-930-7228	Sequence 7228, Ap
42	130.5	8.0	353	US-10-425-114-61294	Sequence 61294, A
43	130.5	8.0	356	US-10-424-599-259801	Sequence 259801,
44	130.5	8.0	366	US-10-424-599-211485	Sequence 211485,
45	130.5	8.0	382	US-10-425-114-66617	Sequence 66617, A

ALIGNMENTS

RESULT 1

US-09-924-841-18
Sequence 18, Application US/09924841
Patent No. US20020127633A1

GENERAL INFORMATION:

APPLICANT: Dilley, David R
APPLICANT: Kadyrzhanova, Dina K
APPLICANT: Wang, Zhenyong
APPLICANT: Warner, Toni M
TITLE OF INVENTION: Modified Synthetases To Produce Penicillins and
TITLE OF INVENTION: Cephalosporins Under the Control of Bicarbonate
FILE REFERENCE: MSU41-453
CURRENT APPLICATION NUMBER: US/09/924, 841
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US/09/413, 231
PRIOR FILING DATE: 1999-10-06
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 18

LENGTH: 311
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: modified DAOCS
OTHER INFORMATION: from Streptomyces clavuligerus
NAME/KEY: MUTAGEN
LOCATION: (181)
OTHER INFORMATION: Ala181 in native DAOCS modified to Arg

Query Match 99.6%; Score 1629; DB 9; Length 311;
Best Local Similarity 99.4%; Pred. No. 1.9e-16;
Matches 309; Conservative 1; Mismatches 1; Gaps 0;

```

QY      1 MDTVPTESLAELOOGLHODEFRRCRLDKGLFYLTDGCLTDTTELKSAKDVIDFEEHSGSE 60
D      1 MDTVPTESLAELOOGLHODEFRRCRLDKGLFYLTDGCLTDTTELKSAKDVIDFEEHSGSE 60
QY      61 AEKRAVTSPTVMRGTGLESSESTAOITNTGYSIDYSCMYSMTADNLPFSGDFERIMT 120
D      61 AEKRAVTSPTVMRGTGLESSESTAOITNTGYSIDYSCMYSMTADNLPFSGDFERIMT 120
QY      121 QYPRORTASPAVAAREVLRATGTETPDGVEAFLLDCEPLLRFRYPOVPEHRSABEQPLRM 180
D      121 QYPRORTASPAVAAREVLRATGTETPDGVEAFLLDCEPLLRFRYPOVPEHRSABEQPLRM 180
QY      181 APHYDLSVTLTIQOTPCANGFVSLOAEVGAFTDLPYRPDAVLVFCGAIATLVGGGVKA 240
D      181 APHYDLSVTLTIQOTPCANGFVSLOAEVGAFTDLPYRPDAVLVFCGAIATLVGGGVKA 240
QY      241 PRHHVAAPRRDQIAGSSRTSSVFLRPVADPTFSPVPLARECGFVSLDGETATFQDWIGG 300
D      241 PRHHVAAPRRDQIAGSSRTSSVFLRPVADPTFSPVPLARECGFVSLDGETATFQDWIGG 300
QY      301 NYVNIIRRTSKA 311
D      301 NYVNIIRRTSKA 311

```

RESULT 2

```

US-10-732-923-9965
; Sequence 9965, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 9965
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Rhodospirillum rubrum
US-10-732-923-9965

```

Query Match 10.6%; Score 173.5; DB 17; Length 345;
 Best Local Similarity 26.5%; Pred. No. 1.6e-09;
 Matches 80; Conservative 33; Mismatches 134; Indels 55; Gaps 12;

```

QY      6 PTFSL-AELOOGLHODEF---RRCLRDKGLFYLTDGCLTDTTELKSAKDVIDFEEHSGSE 60
D      13 PVLSLKAGEGAARDEFIATLRQARDPGAFLEHGCHIDPSVIDQVEALSRRFFAL-PE 71
QY      61 AEKRAVTSPTVMRGTGLESSESTAOITNTGYSIDYSCMYSMTADNLPFSGDFERIMT 120
D      72 DEKRAIDMVNPPHFGYTRVGAEILTR-----GAPDMREQDIDGSEBRLPQGPDTAWA 125
QY      121 QY-----FDRQYTA-SRAVAREVLRATGTETPDGVEAFLLDCEPLLRFRYPOVPEHRSABEQPLRM 155
D      126 RLQGNQWPAALPDLRAAVLRILQAELETVLALALERIALALAGERAD---FFADLYEGGP 181
QY      156 EPLLRFRYPOVPEHRSABEQPLRMAPHYDLSMTLLIQOTPCANGFVSLOAEVGAFTDLPYRPDAVL 215
D      182 DQLKIRYF---GRAGBEGDQGVPHKDSGLLTFVLQRRG---GLQIEREGHWVDV 233
QY      216 PYRPDAVLVFCGAIATLVGGGVKAPRHHVAAPRRDQIAGSSRTSSVFLRPVADPTFSPVPLARECGFVSLDGETATFQDWIGG 275
D      234 PPRGTFAVNIIGELLELATNTGVLKATYHVVSPAD-----SDRLSIAFFL--GARLSKV 287
QY      276 PL 277
D      288 PL 289

```

```

RESULT 3
US-10-732-923-9978
; Sequence 9978, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 9978
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Magnetospirillum magnetotacticum
US-10-732-923-9978

```

Query Match 10.5%; Score 171; DB 17; Length 342;
 Best Local Similarity 26.3%; Pred. No. 3e-09;
 Matches 87; Conservative 38; Mismatches 154; Indels 52; Gaps 14;

```

QY      8 FSLAELOOGLHODEFRRCRLDKGLFYLTDGCLTDTTELKSAKDVIDFEEHSGSEAEKRAVT 67
D      16 FDGDAERAPFLDDLRAARDVGFYLTGHGLPTLELAIALQALARRFFAL-PLPEGRVA 74
QY      68 SPVPTMRGTGLESSESTAOITNTGYSIDYSCMYSMTADNLPFSGDFERIMTQYPRDQ- 126
D      75 MVASPHFRGYT---BAGREIYR--GRPDWREQDAGABERTVLPWHPGLPAMTRLOGPNU 128
QY      127 -----YTASRAVAREVLRATGTETPDGVEAFLLDCEPLLRFRYPOVPEHRSABEQPLRM 163
D      129 WPSLDLRTGLVAKQEAATGTGIRLLRAPALSLOPPADP---EPIYGAEPNQLKTIIR 185
QY      164 FPVPEHRSABEQPLRMAPHYDLSMTLLIQOTPCANGFVSLOAEVGAFTDLPYRPDAVL 223
D      186 YP-----GRDAERDAQGVGHKDSGFLITLLQD--GGGLEVEGEDGWTAAAPVE-GAVY 238
QY      224 VFCGAIATLVGGGVKAPRHHVAAPRRDQIAGSSRTSSVFLRPVADPTFSPVPLARECGFVSLDGETATFQDWIGG 275
D      239 VNVGELLEIASNGLFRAIVHVVAP---AAGRDLISVAFLEGARHARDATVPPLSLPPDLA 294
QY      279 RECQFVSLDGETATFQDWIGGNYVNIIRRTS 309
D      295 AQARGPAS-DPDNPLFRE-VGRNYLKGRLRS 323

```

RESULT 4

```

US-10-369-493-12239
; Sequence 12239, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianteng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12239
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-12239

```

Query Match 10.3%; Score 169; DB 15; Length 366;

Best Local Similarity 24.7%; Pred. No. 5.4e-09;
Matches 85; Conservative 41; Mismatches 158; Indels 60; Gaps 14;

```
QY 1 MDTTPTFSLAIQGLHQ-----DEFRCLNDKGLFYITDGLDTELSKADVIDDF 55
DB 28 MPRIVPVLDSLRLSEQASERRTFPLDLRSASRDIGFYLAGHGISWAEISVLTASRQPF 87
QY 56 EHGSSEKRAVTSVPVTRMGFTGLESESTAQITNTGSDYSMCSMGTANLFPSSGDF 115
DB 88 AL-PEADKLAIEWYSSQFRGTYRAGGELTK-----GREDMREQLDIVEROAIAGG 140
QY 116 ERIWTO-YFDROYTAS-----RAVAEVLNATGTEPDGVEAFLDCEPFL 159
DB 141 TPAMTRLOGPNOGMPALPDLKALLAWOSKVTAVAIRLKAFAQSLDQPEDAF---DPI- 196
QY 160 RRRYFQVDEHR-SAEBOPLR-----MAPHYDISMTTLIQPTPCANGFVSLQAEVGG 210
DB 197 ---YSSBPNHMKIVRYRPGRDITGGDQGVGAHKDGGFTLLLO---DNNKGLDQVDYDG 248
QY 211 APTDLPYRDAVLVEFGATATLVTTGGQVAPRHHVAAPRRDQIAGSSRTSVFELRPND 270
DB 249 SWVDVDPITGTLVNIIGELLEIASNGYLAIVYHVVTP---AGVERISVPEFFSARLD 304
QY 271 FT-----FSVPLARECGFDVSLDGETATFQDWIGGNYVNIIRTS 309
DB 305 ATIPILGLSEELAQAARGPAS-DPDNPLFRD-VGTNVLSKRLRS 346
```

RESULT 5
US-10-732-923-10056
; Sequence 10056, Application US/10732923
; Publication No. US20050108791A1

GENERAL INFORMATION:
; APPLICANT: Edgeton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 10056
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-732-923-10056

Query Match 10.3%; Score 169; DB 17; Length 366;
Best Local Similarity 24.7%; Pred. No. 5.4e-09;
Matches 85; Conservative 41; Mismatches 158; Indels 60; Gaps 14;

```
QY 1 MDTTPTFSLAIQGLHQ-----DEFRCLNDKGLFYITDGLDTELSKADVIDDF 55
DB 28 MPRIVPVLDSLRLSEQASERRTFPLDLRSASRDIGFYLAGHGISWAEISVLTASRQPF 87
QY 56 EHGSSEKRAVTSVPVTRMGFTGLESESTAQITNTGSDYSMCSMGTANLFPSSGDF 115
DB 88 AL-PEADKLAIEWYSSQFRGTYRAGGELTK-----GREDMREQLDIVEROAIAGG 140
QY 116 ERIWTO-YFDROYTAS-----RAVAEVLNATGTEPDGVEAFLDCEPFL 159
DB 141 TPAMTRLOGPNOGMPALPDLKALLAWOSKVTAVAIRLKAFAQSLDQPEDAF---DPI- 196
QY 160 RRRYFQVDEHR-SAEBOPLR-----MAPHYDISMTTLIQPTPCANGFVSLQAEVGG 210
DB 197 ---YSSBPNHMKIVRYRPGRDITGGDQGVGAHKDGGFTLLLO---DNNKGLDQVDYDG 248
QY 211 APTDLPYRDAVLVEFGATATLVTTGGQVAPRHHVAAPRRDQIAGSSRTSVFELRPND 270
DB 249 SWVDVDPITGTLVNIIGELLEIASNGYLAIVYHVVTP---AGVERISVPEFFSARLD 304
QY 271 FT-----FSVPLARECGFDVSLDGETATFQDWIGGNYVNIIRTS 309
```

DB 305 ATIPILGLSEELAQAARGPAS-DPDNPLFRD-VGTNVLSKRLRS 346

```
RESULT 6
US-09-924-841-5
; Sequence 5, Application US/09924841
; Patent No. US20020127633A1
GENERAL INFORMATION:
; APPLICANT: Dilley, David R
; APPLICANT: Kadyrzhanova, Dina K
; APPLICANT: Wang, Zhenyong
; APPLICANT: Warner, Toni M
; TITLE OF INVENTION: Modified Synthetases To Produce Penicillins and
; FILE REFERENCE: MS041-453
; CURRENT APPLICATION NUMBER: US/09/924,841
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US/09/413,231
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Artificial Sequence
FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: modified IPNS
; OTHER INFORMATION: from Streptomyces clavuligerus
; NAME/KEY: MUTAGEN
; LOCATION: (210)
; OTHER INFORMATION: Glu210 in native IPNS modified to Arg
US-09-924-841-5
```

Query Match 10.3%; Score 168; DB 9; Length 329;
Best Local Similarity 23.8%; Pred. No. 6e-09;
Matches 77; Conservative 50; Mismatches 150; Indels 46; Gaps 14;

```
QY 5 VPTFSIABL-----QCSLHODEFRRCIRDKGLFYITDGLDTELSKADVIDDF 58
DB 10 VPTDISPLFTDAAAKRVAEIHGACRSGGFYATNHG---DVQQLQDVNVEFGAM 66
QY 59 SEAEK-----RAVTSVPVTRMGF-----TGLESESTAQITNTGSDYSMCSMGT---A 106
DB 67 TDQEKHDAIAHAYPNPDNPHVANGYKAPGKAVSPCYLNPDPGEHPRM-IAAGTRPHE 125
QY 107 DNLFPSSGD---FERIWTQYDROYTASRAVAEVLNATGTEPDGVEA-----FLDCE 156
DB 126 VNLMPDEERHHRFPFCGYYRQMLKSTVLMRGIALALG-RPHHFPDAAALAEODSLSSV 184
QY 157 PLLEPRRYPOVPEHRSAAE-OLPMAAPHYDISMTTLIQPTPCANGFVSLQAEVGAFTDL 215
DB 185 SLIRPYEERYPVKCTGPDGQLSFRDHLDSMTVLEFQTVON---LOVEYDGMWDI 240
QY 216 PYRDAVLVEFGATATLVTTGGQVAPRHHVAAPRRDQIAGSSRTSVFELRPND 275
DB 241 PTSENDPLVNGCYTMAHNTNDYFPAPNHRV-----KVNARLSLPFFLNGHEAVYE- 293
QY 276 PLARECGFDVSLDGETATFQDWI 298
DB 294 FVPE-GESEVRNEALSVDYL 315
```

RESULT 7
US-09-924-841-9
; Sequence 9, Application US/09924841
; Patent No. US20020127633A1
GENERAL INFORMATION:
; APPLICANT: Dilley, David R
; APPLICANT: Kadyrzhanova, Dina K
; APPLICANT: Wang, Zhenyong
; APPLICANT: Warner, Toni M
; TITLE OF INVENTION: Modified Synthetases To Produce Penicillins and
; Cephalosporins Under the Control of Bicarbonate

Query Match	10.3%	Score 168;	DB 9;	Length 329;
Best Local Similarly	23.8%	Pred. No. 6e-09;		
Matches 77; Conservative	50;	Mismatches 150;	Indels 46;	Gaps 14;

```

RESULT 8
US-10-732-9964
: Sequence 9964, Application US/10732923
: Publication No. US20050108791A1
: GENERAL INFORMATION:
: APPLICANT: Edgeton, Michael D
: TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
: FILE REFERENCE: 38-15(52796)C
: CURRENT APPLICATION NUMBER: US/10/732,923
: CURRENT FILING DATE: 2003-12-10
: PRIOR APPLICATION NUMBER: 10/7310,154
: PRIOR FILING DATE: 2002-12-04
: NUMBER OF SEQ ID NOS: 24149
: SEQ ID NO 9964
:
: LENGTH: 412
:
: TYPE: PRT
:
: ORGANISM: Rhodospseudomonas palustris
US-10-732-923-9964

```

Db	Sequence	Score	DB	Length
Db	LPVLIDLSFRSDTAEBAEFLFNDVDAAGPGGFFIYVGHGISRLRLRDVLFASRNFPAL-P	130		
Qy	BAEK---RAVYSPVYTMKRGFTGLSESTAOITNTGSYSDYSCYSGMTADNLFPSGDE	116		
Db	EADKLIDIMINSF---HFRGYTRAGREFTR-----CGQDWMEQQLDVGAEREAFLSRSA	181		
Qy	RIWYQYFDRQYTAASAFAREVLRAATGTEBDGVEAFPLDCEPLLRRIYFPQVE-----	169		
Db	PPWTR-----LQGNQWPD-----ALPELKLPL-LRYQOEVELAIKVL	220		
Qy	170-----HRSAEQPLRMAPHYDLSMTLLIOQTPCA	198		
Db	221 VFMAALGQAEDEVFEITYPSPNQLIKIRYFGRADDESQGVGTHKDSGFVITLLDPTVA	280		
Qy	199 NGFVSLQAEVGAFYDLPYRPAVLVFCGAIATLVYTGQVKAPRHHVAAPREDQIAGSSR	258		
Db	281 ---GLQYETADGWIDAPPLPGSFVNNIGELIELSNGALRANVHRVSPDPD---TDR	332		
Qy	259 TSSVFFLRPNADFTSY---PLAECCGPDVSLDGETATFQWIGANTYINIRTS	309		
Db	333 LSAFFLGAFLRDLATPVVLTLPPELADYARGVYQDPQNPFLRD-VGRNNYLKGRILRS	386		
RESULT 9				
US-10-425-115-220323				
; Sequence 220323, Application US/10425115				
; Publication No. US20040214272A1				
; GENERAL INFORMATION:				
; APPLICANT: La Rosa, Thomas J.				
; APPLICANT: Kovalic, David K.				
; APPLICANT: Zhou, Yihua				
; APPLICANT: Cao, Yongwei				
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with				
; TITLE OF INVENTION: Plants				
; FILE REFERENCE: 38-21(53222)B				
; CURRENT APPLICATION NUMBER: US/10/425,115				
; CURRENT FILING DATE: 2003-04-28				
; NUMBER OF SEQ ID NOS: 369326				
; SEQ ID NO 220323				
; LENGTH: 342				
; TYPE: PRT				
; ORGANISM: Zea mays				
; FEATURE:				
; OTHER INFORMATION: Clone ID: MRT4577_132519C.1.pep				
US-10-425-115-220323				
Query Match				
Best Local Similarity 23.4%; Pred. No. 1.2e-07;				
Matches 78; Conservative 49; Mismatches 119; Indels 88; Gaps 16;				
Qy	4 TVPTSLAELOQGLHODEFRCL-----RDKGLFYLTCGLTDTLKSARDIVDFEHGS	59		
Db	31 TSATVSLPIVDSLDRRAILAEKKEIGFQVNVHGSLEAMQDMETVCGEPR--L	88		
Qy	60 EAERAAVTSPVYTMKRGFTGLSESTAOIT-----NTGSYSDYSCYSGMTADNLFP	111		
Db	89 PAEDRA-----GLYSEDTGRATRIYSSTMPDTGAEKWRDCLRLACS---	132		
Qy	112 S-GDPERIT-----TOYFDRQYTAASRAVAREVL-----ANGTEPD-----GGEVAF	152		
Db	133 AAGSDAAMPDKPRRLREVERFTVOTRGLKEIHLRLCEGLGRDYEIGDISG----	188		
Qy	153 LDCEBILRFERYPOVEHRSABEOPLRMAPHYDLSMTLLIOQTPCANGFVSLQAEVGAF	212		
Db	189 ---DVVLVHNHPRCPDPRAT-----LGLPRHCDRLNLTLL-----LPSWVG--	228		
Qy	213 TDLYR-----PDVAVLFCGAIATLVYTGQVKAAPRHHVAAPRROQIAGSSRSVF	263		
Db	229 LEAARGDWIRVEPVGAFVNVFGCOLLEVITNGILKSIHRVMTN-----LGVARTTVAT	283		
Qy	264 FLRPNADFTFSVPLARECGFDVSLDGETATFQDW	297		
Db	284 FIMPTTDLIG-PAAEFLSDNNPCRYTITLADF	316		

```
RESULT 10
US-10-425-114-58493
; Sequence 58493, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 58493
; LENGTH: 375
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3587-225-H12_FLI pep
US-10-425-114-58493
```

```
Query Match          9.5%; Score 156; DB 15; Length 375;
Best Local Similarity 23.4%; Pred. No. 1.4e-07;
Matches 78; Conservative 49; Mismatches 119; Indels 88; Gaps 16;

QY 4 TPTFSLAELOOGLHODEFRRCI-----RDKGLFYLTDCGLTDTBELKSAKDIIVDFEBSGS 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 64 TSATVSLPIVDLSLGRDVRRAILEAGKEIGFPQVNHGVSLEAMQMETVCOEFR--L 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 60 EAERKAVTSPVPTMRGFTGLESESTAOIT-----NTGSYSDYSNMGTDNLFP 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 122 PAEDKA-----GLYSEDYGRATRIYSTMPDTGAEKWRCRLACS---FP 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 112 S-GDPERIW-----TOYFRQYTAARAAREVLR---ATGTEPD-----GCVNAF 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 166 AVGDSAAAMPDKPRRLREVERFTVQTRGLGMEILRLCEGLGRPDYLEGDISG--- 221
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 153 LDCEPLLRFRYPPOVPEHRSABEQLRMAPHYDLSMTLTIQOTPCANGFVSLQAEVGAF 212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 222 ---DVLVHNHYPCPDPMAT---LGLPFCRNLTL-----LPSWVPG-- 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 213 TDLPYR-----PDAVLVFCGAIATLVYTGQVAKPRHHVAAPRRDQIAGSSRTSSVF 263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 262 LEVAVRGDMIRVEPVGAFVNVFGQLLEVVTNGILKSIHRVMTN-----LGVARTVAT 316
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 264 FLRRNADFTFSVPLARECGFVSLDGETATPDW 297
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 317 FIMPTTDLIG-PAAEFLSDDNPPCYRTLTFADF 349
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 11
US-10-425-114-63175
; Sequence 63175, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
```

```
; SEQ ID NO 63175
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3062-030-F8_FLI pep
US-10-425-114-63175
```

```
Query Match          9.5%; Score 155; DB 15; Length 352;
Best Local Similarity 23.4%; Pred. No. 1.7e-07;
Matches 78; Conservative 49; Mismatches 119; Indels 88; Gaps 16;

QY 4 TPTFSLAELOOGLHODEFRRCI-----RDKGLFYLTDCGLTDTBELKSAKDIIVDFEBSGS 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 41 TSATVSLPIVDLSLGRDVRRAILEAGKEIGFPQVNHGVSLEAMQMETVCOEFR--L 98
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 60 EAERKAVTSPVPTMRGFTGLESESTAOIT-----NTGSYSDYSNMGTDNLFP 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99 PAEDKA-----GLYSEDYGRATRIYSTMPDTGAEKWRCRLACS---FP 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 112 S-GDPERIW-----TOYFRQYTAARAAREVLR---ATGTEPD-----GCVNAF 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 143 AVGDSAAAMPDKPRRLREVERFTVQTRGLGMEILRLCEGLGRPDYLEGDISG--- 198
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 153 LDCEPLLRFRYPPOVPEHRSABEQLRMAPHYDLSMTLTIQOTPCANGFVSLQAEVGAF 212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 199 ---DVLVHNHYPCPDPMAT---LGLPFCRNLTL-----LPSWVPG-- 238
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 213 TDLPYR-----PDAVLVFCGAIATLVYTGQVAKPRHHVAAPRRDQIAGSSRTSSVF 263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 239 LEVAVRGDMIRVEPVGAFVNVFGQLLEVVTNGILKSIHRVMTN-----LGVARTVAT 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 264 FLRRNADFTFSVPLARECGFVSLDGETATPDW 297
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 294 FIMPTTDLIG-PAAEFLSDDNPPCYRTLTFGDF 326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 12
US-10-437-963-146275
; Sequence 146275, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Bouharov, Andrej A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 146275
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_46914C.1 pep
US-10-437-963-146275
```

```
Query Match          9.4%; Score 154.5; DB 16; Length 366;
Best Local Similarity 22.1%; Pred. No. 2e-07;
Matches 70; Conservative 52; Mismatches 158; Indels 37; Gaps 10;

QY 5 VPTFSLAELOOGLHQ--DEPRRCIRDKGLFYLTDCGLTDTBELKSAKDIIVDFEBSGAE 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 64 IPDIVGELQGSDELIDNRLACEQWGFQVNVHGVSEETMEKARAEFFFLPLEEK 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 63 KRAVTSVPPTMRGFTGLESESTAOITNTGSYSDYSNMGTD-----ADNLPPS--GD 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db      124 EKVMEP-----GGIGGHAFFVSDDDOKLDMCNMLAGVBPATIRRNLMPTTPAN 175
      115 FERLWTOYFROYTASRAVAEBVLATPTEPDGGVEAFLDCEPLLRFERYPOVEHRSAE 174
      176 FSKTLEKYSVEIRELCVRLHEHIAAALGLAPRLNGMGEGVAQVQVRNMFPPCR-----P 231
      175 EQLPMAPHYLSWVTLIQOTPCANGFVSLQAEVGAFDTLPRPADVLVFCGAMATLV 234
      232 ELVLGLSHSDGSAVTVLQODAAFPAGLOVLRG--GGGVAAVAVPVGALVNAVGDITLEVLT 289
      235 GGOVAPPRHH-VAAPRRDQIAGSSRTSSVFLRPVADFTFSVPLARECGFDVSLDGETAT 293
      290 NGRKYSVHRAVASEHD-----RMSVTVFAPAVDVLC-PLP-----ELVADGERRR 337
      294 FQDWIGYVAVIRRTSK 310
      338 YRTVMGEYSRHHVYTSR 354
```

RESULT 13

```
US-10-425-115-220322
; Sequence 220322, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 220322
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_132518C.1.pcp
US-10-425-115-220322
```

```
Query Match      9.4%; Score 154; DB 16; Length 342;
Best Local Similarity 23.4%; Pred. No. 2e-07;
Matches 78; Conservative 48; Mismatches 120; Indels 88; Gaps 16;
```

```
      4 TVPTFSLAELQGLHODEFRRL-----RDKGLFYLTDCGLTDTLKSAXDIVIDFEHGS 59
      31 TSAFVSLPIVDLSIGRDEVRRAILEAGKEIGFQVNVHGVSLAMQDMETVCGEFPFR--L 88
      60 EAEKRAVTSVPVPMRGRGTGSESTAOIT-----NTGSYDSYSCYMGATADNLFP 111
      89 PAEDKA-----GLYSEDTGRATRIYSTMTFDGEEKYWRDCLRLACS---FP 132
      112 S-GDPERIW-----TOYFDRQYASRAVAREVLR---ATGTEPD-----GGVEAF 152
      133 AVGDSAAAMPDKRRLREVERFTYOTRGIMELIRLCEGLGRPDVLEBDISGG----- 188
      153 LDCEPLIRFRYPPOVPEHRSABEQPLRMAPHYDLSWVTLIQOTPCANGFVSLQAEVGAF 212
      189 ---DVVLHVNHYPCCDPNAT---LGLPPHCDRNLTLT-----LPSMVPG-- 228
      213 TDLPYR-----PDAVLVFCGAIATLVYGGQVAKPRHHVAAARRDQIAGSSRTSSVF 263
      229 LEVARRGDMIRVBPVGAFVNFQGLEEVVTVNGILKSIHRVMTN-----LGVARTTVAT 283
      264 FLRPNADFTFSVPLARECGFDVSLDGETATFQDW 297
      284 FIMPTTDLCLG-PAAEFLSDNDNPPCYRTLITFGDF 316
```

RESULT 14

```
US-10-425-114-62424
; Sequence 62424, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaka, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 62424
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3357-058-F2_FLI.pcp
US-10-425-114-62424
```

```
Query Match      9.4%; Score 154; DB 15; Length 371;
Best Local Similarity 23.4%; Pred. No. 2.3e-07;
Matches 78; Conservative 48; Mismatches 120; Indels 88; Gaps 16;
```

```
      4 TVPTFSLAELQGLHODEFRRL-----RDKGLFYLTDCGLTDTLKSAXDIVIDFEHGS 59
      60 TSAFVSLPIVDLSIGRDEVRRAILEAGKEIGFQVNVHGVSLAMQDMETVCGEFPFR--L 117
      60 EAEKRAVTSVPVPMRGRGTGSESTAOIT-----NTGSYDSYSCYMGATADNLFP 111
      118 PAEDKA-----GLYSEDTGRATRIYSTMTFDGEEKYWRDCLRLACS---FP 161
      112 S-GDPERIW-----TOYFDRQYASRAVAREVLR---ATGTEPD-----GGVEAF 152
      162 AVGDSAAAMPDKRRLREVERFTYOTRGIMELIRLCEGLGRPDVLEBDISGG----- 217
      153 LDCEPLIRFRYPPOVPEHRSABEQPLRMAPHYDLSWVTLIQOTPCANGFVSLQAEVGAF 212
      218 ---DVVLHVNHYPCCDPNAT---LGLPPHCDRNLTLT-----LPSMVPG-- 257
      213 TDLPYR-----PDAVLVFCGAIATLVYGGQVAKPRHHVAAARRDQIAGSSRTSSVF 263
      258 LEVARRGDMIRVBPVGAFVNFQGLEEVVTVNGILKSIHRVMTN-----LGVARTTVAT 312
      264 FLRPNADFTFSVPLARECGFDVSLDGETATFQDW 297
      313 FIMPTTDLCLG-PAAEFLSDNDNPPCYRTLITFGDF 345
```

```
RESULT 15
US-10-425-114-53454
; Sequence 53454, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaka, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 53454
; LENGTH: 372
; TYPE: PRT
```

; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700169504_FLI.pep
US-10-425-114-53454

Query Match 9.4%; Score 154; DB 15; Length 372;

Best Local Similarity 23.4%; Pred. No. 2.3e-07;

Matches 78; Conservative 48; Mismatches 120; Indels 88; Gaps 16;

```
QY 4 TVPTSLAEIQGLHODEFRCL---RDKGLPYLTDCGLTDTLKSAXDIVIDPPEHGS 59
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 TSATVSLPIVDLSLGRDEVRAILLEAGKEIGFFQVNVHGVSLAMQDMETVCQEFFR--L 118
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 60 EAEKRAVTSPPVPTMRGFTGLESESTAQIT-----NTGSYSDYSMCYSGTADNLP 111
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 119 PAEDKA-----GLYSBDTGRATRIYSSMTFDTGGEKTYRDCRLIACS--P 162
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 112 S-GDPERIW-----TOYFDRQYASRAVAREVLK---ATGTEPD-----GVEAF 152
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 163 AVGDSAAAMPDKPRRLREVERFTVOTRGLGMEILRLCEGLGRPDYLEGDISGS---- 218
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 153 LDCEPLRRRYFPQVPEHRSABEQPLRNAPHYDLSNVTLIQOTPCANGFVSLQAEVGA 212
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 219 ---DVLHVNHYPPCEDPNAT---LGLPHCDRNILTL-----LPSNVPG-- 258
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 213 TDLPYR-----PDAYLVFCGAIALTVTGGOVKAPRRHHVAPRRDQIAGSRTSSVF 263
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 259 LEAVTRGDMIRVEVPVPGAFVNVFGCOLEVVINGILKSIHRVMTN----LGVARTVAT 313
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 264 FLRPNADFTFSVPLARECGFDVSLDGETATFQDW 297
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 314 FIMPTTDCIIG--PAAEPLSDDNPPCYRTLTTFGDF 346
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: September 19, 2005, 15:24:23
Job time : 86 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 19, 2005, 15:19:28 ; Search time 4917 Seconds

(without alignments)
3064.790 Million cell updates/sec

Title: 10719236-1_232-1164

Perfect score: 1636

Sequence: 1 MDTVPTFTSLAELOQGHQD.....ATFQDWIGNVYINIRTSKA 311

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=xlh
-O=/g2n.1/USPTO_spool/h/10719236-1TRANS/runat_19092005_153157_7740/app_query.fasta_1.45
-DB=GenEmbl -QFMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -STRAT=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi -LIST=45
-DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=10719236-1TRANS @CGN_1_1_3731_@runat_19092005_153157_7740 -NCPUL=6
-ICPU=3 -NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_srs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1636	100.0	1230	6	107104 Sequence 1
2	1636	100.0	1230	6	AR479090 Sequence
3	1634	99.9	936	6	AX300747 Sequence
4	1634	99.9	2494	1	STWCERDA M32324 S.clavulig

5	1385	84.7	936	1	AY318742
6	1378	84.2	1141	1	AF317908
7	1300	79.5	936	1	AY318743
8	1210	74.0	939	6	A42959
9	1210	74.0	939	6	A72870
10	1210	74.0	939	6	AR023763
11	1210	74.0	939	6	AR023763
12	1195.5	73.1	942	6	A43001
13	1195.5	73.1	942	6	A72872
14	1195.5	73.1	942	6	AR023762
15	1195.5	73.1	942	6	191795
16	1195.5	73.1	942	6	NLCERB
17	1195.5	73.1	2169	1	SS5763
18	1195.5	73.1	2169	1	AY318744
19	1012	61.9	939	1	NLCERB
20	972.5	59.4	933	1	NLCERB
21	946.5	57.9	954	1	STWCERF
22	944	57.7	6358	1	ILYK90C
23	933.5	57.1	1511	6	105489
24	924.5	56.5	5775	8	CQ796893
25	427	26.1	5960	1	ACH404737
26	183	11.2	1044	1	SCCERGEN
27	180	11.0	346357	1	STWIPNSSJ
28	178	10.9	990	1	SGPCBC
29	169	10.3	340857	1	AP003010
30	168	10.3	990	1	STWIPNS
31	168	10.3	990	6	A01132
32	168	10.3	29870	6	AX763606
33	168	10.3	349640	1	BE016866
34	166.5	10.2	310029	1	AF320779
35	165	10.1	990	1	AR029451
36	165	10.1	1250	6	AR029451
37	164	10.0	1310	1	FSPCBC
38	163	10.0	786	1	AB045851
39	161	9.8	36687	1	AY260760
40	160	9.8	786	1	AB045853
41	158	9.7	786	1	AB045852
42	158	9.7	786	1	AB045857
43	158	9.7	1331	8	BT013114
44	156.5	9.6	1891	3	AT075378
45	156.5	9.6	1938	3	AY102699

ALIGNMENTS

RESULT 1					
107104	Sequence 1	from Patent EP 0341892.	1230 bp	DNA	linear
LOCUS	107104				
DEFINITION	Sequence 1	from Patent EP 0341892.			
ACCESSION	107104				
VERSION	107104.1	GI:589821			
KEYWORDS					
SOURCE					
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 1230)				
AUTHORS	Ingolia,T.D., Kovacevic,S., Miller,J.R. and Skatrud,P.L.				
TITLE	Recombinant DNA expression vectors and DNA compounds that encode				
JOURNAL	deacetoxycephalosporin C synthetase				
FEATURES	Patent: EP 0341892-A1 1 15-NOV-1989;				
source	location/Qualifiers				
	1..1230				
	/organism="unknown"				
	/mol_type="unassigned DNA"				

ORIGIN

Alignment Scores:	2.7e-136	Length:	1230
Pred. No.:	1636.00	Matches:	311
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	100.00%	Gaps:	0
DB:	6		

10719236-1_232-1164 (1-311) x 107104 (1-1230)

QY 1 MetAspThrThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAsp 20
Db ATGACACGACGAGCGTGCCCACTTCAGCTGGCCGCAACTCCAGCGGGCTTCGACCAAGAC 291

QY 21 GluPheArgArgCysLeuArgAspLysGlyLeuPheTyrLeuThrAspCysGlyLeuThr 40
Db GAGTTCGGCAGGTGTCTGAGGAGCAAGGGCTCTTCTATCTGACGAGCTGGGTTCTAGCC 351

QY 41 AspThrGluLeuLysSerAlaLysAspIleValIleAspPhePheGlnHisGlySerGlu 60
Db GACACCGAGCTGAGCTGGCCCAAGCAATCTCTCATTCGACTTCTCGAGCAGCGAGGAG 411

QY 61 AlaGluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeu 80
Db GCGGAGAAAGCCGCGGTCACTCCCGTCCCACTGCGCCGCGGCTTCACCGGGCTG 471

QY 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys 100
Db GAGTCGAGAGACACCGCCAGATCACCAATACCGGAGCTACTCCGACTCATGTGTC 531

QY 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTyrThr 120
Db TACTCGATGGGCAACCGGAGCAACCTCTCCGTCGCGTGCCTTCGAGCGGATCTGAGAC 591

QY 121 GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
Db CAGTACTTCGACCCCGAGTACACCGCTCCCGCGGTGCGCCGCGGAGGTCTTCGCGGCG 651

QY 141 ThrGlyThrGluProAspGlyGlyValGluValaPheLeuAspCysGluProLeuLeuArg 160
Db ACCGGAGCCGAGCCCGAGCGGGGTGAGGCTTCTCTGACTCGAGCCGCTCTGCGG 711

QY 161 PheArgTyrPheProGlnValProGlnHisArgSerAlaGluGlnGlnProLeuArgMet 180
Db TTCGGTACTCTCCGCAAGGTCCCGAGACCGGACCGCGGAGCGGAGCGCCCTGCGAGT 771

QY 181 AlaProHisTyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGly 200
Db GCGCGGACTACGACCTGTGATGTGACCTCTCATCCGACGACCTGCTTACCTCGGAC 831

QY 201 PheValSerLeuGlnAlaGluValGlyValaPheThrAspLeuProTyrArgProAsp 220
Db TTGCTGACCTTCAGGCGGAGGTGCGCGCGCTTCACGAGCTGCGCTTACCTCGGAC 891

QY 221 AlaValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyGlnValLysAla 240
Db GCCGTCCGTCTCTGCGGCGCATGCGACCTGTGAGCGGCGGCGGAGGTCAAGGCC 951

QY 241 ProArgHisHisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260
Db CCCCCGACCACTGTGCGCGCCCCCGGAGGAGCCAGATAGCGGCGGAGCGGACCTCC 1011

QY 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280
Db AGTGTTCT 1071

QY 281 CysGlyPheAspValSerLeuAspGlyGlnThrAlaThrPheGlnAspTyrIleGlyGly 300
Db TGGCGGCTTCGATGTGAGCTGAGCGGAGACCGGACGTTCCAGGATTTGATGAGCGGG 1131

QY 301 AsnTyrValaLeuIleArgArgThrSerLysAla 311
Db AACTACGTGAACATCCGCGCACATCCAAAGGCA 1164

RESULT 2
AR479090 1230 bp DNA linear PAT 14-MAY-2004
DEFINITION Sequence 1 from patent US 6699699.
ACCESSION AR479090
VERSION AR479090.1 GI:47238046

KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1230)
AUTHORS Yunn-Bor,Y., Chia-Li,W., Jyh-Shing,H. and Ying-Chieh,T.
TITLE Mutated penicillin expandases
JOURNAL Patent: US 6699699-A 1 02-MAR-2004;
FEATURES
source
location/Qualifiers
1..1230
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 2,7e-136 Length: 1230
Score: 1636.00 Matches: 311
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

10719236-1_232-1164 (1-311) x AR479090 (1-1230)

QY 1 MetAspThrThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAsp 20
Db ATGACACGACGAGTGCCCACTTCAGCTGGCCGCAACTCCAGCGGGCTTCGACCAAGAC 291

QY 21 GluPheArgArgCysLeuArgAspLysGlyLeuPheTyrLeuThrAspCysGlyLeuThr 40
Db GAGTTCGGCAGGTGTCTGAGGAGCAAGGGCTCTTCTATCTGACGAGCTGGGTTCTAGCC 351

QY 41 AspThrGluLeuLysSerAlaLysAspIleValIleAspPhePheGlnHisGlySerGlu 60
Db GACACCGAGCTGAGCTGGCCCAAGCAATCTCTCATTCGACTTCTCGAGCAGCGAGGAG 411

QY 61 AlaGluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeu 80
Db GCGGAGAAAGCGCGGTCACTCCCGTCCCACTGCGCCGCGGCTTCACCGGGCTG 471

QY 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys 100
Db GAGTCGAGAGACACCGCCAGATCACCAATACCGGAGCTACTCCGACTCATGTGTC 531

QY 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTyrThr 120
Db TACTCGATGGGCAACCGGAGCAACCTCTCCGTCGCGTGCCTTCGAGCGGATCTGAGAC 591

QY 121 GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
Db CAGTACTTCGACCCCGAGTACACCGCTCCCGCGGTGCGCCGCGGAGGTCTTCGCGGCG 651

QY 141 ThrGlyThrGluProAspGlyGlyValGluValaPheLeuAspCysGluProLeuLeuArg 160
Db ACCGGAGCCGAGCCCGAGCGGGGTGAGGCTTCTCTGACTCGACGCGGCTGCGCGG 711

QY 161 PheArgTyrPheProGlnValProGlnHisArgSerAlaGluGlnGlnProLeuArgMet 180
Db TTCGGTACTCTCCGCAAGGTCCCGAGACCGGAGCGCGGAGGAGCGCCCTGCGAGT 771

QY 181 AlaProHisTyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGly 200
Db TGGCGGCTTCGATGTGAGCTGAGCGGAGACCGGACGTTCCAGGATTTGATGAGCGGG 831

QY 201 PheValSerLeuGlnAlaGluValGlyValaPheThrAspLeuProTyrArgProAsp 220
Db TTCGTGAGCTTCAGGCGGAGGTGCGCGGCTTCATCCAGGACCTGCGCTTACCTCGGAG 891

QY 221 AlaValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyGlnValLysAla 240
Db GCCGTCCGTCT 951

QY 241 ProArgHisHisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260

Db 952 CCCCCGACATGTGCGGGCCCCCGGACGAGCACGATGAGGGGAGGAGCGGACCTCC 1011
Qy 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280
Db 1012 AGTGTGTTCTTCTCCGTCCCAACCGCGACTTCTCTCCGTCCGCTGCGCGCGAG 1071
Qy 281 CysGlyPheAspValSerLeuAspGlyValThrAlaThrPheGlnAspTrpIleGlyGly 300
Db 1072 TCGGCTTCGATGTCAAGCTGAGCGGAGCGGACCGGACCTTCAGAGATTGATCGGAGGC 1131
Qy 301 AsnTyValAsnIleArgArgThrSerIlySala 311
Db 1132 AACTACGTGAACATCGCGGACATCCAGGCA 1164
RESULT 3
AX300747 936 bp DNA linear PAT 30-NOV-2001
LOCUS Sequence 1 from Patent WO0185951.
DEFINITION AX300747
ACCESSION AX300747
VERSION AX300747.1 GI:17382047
KEYWORDS
SOURCE
ORGANISM Streptomyces clavuligerus
Bacterium; Streptomyces clavuligerus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE
1 Johnson, R. I. and Newbert, R. W.
A modified expandase and uses thereof
Patent: WO 0185951-A 1 15-NOV-2001;
Acs Dobfar UK Limited (GB)
FEATURES
source location/Qualifiers
1..936
/organism="Streptomyces clavuligerus"
/mol_type="unassigned DNA"
/db_xref="taxon:1901"
1..936
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD13038.1"
/db_xref="GI:17382048"
/translation="MDTVPTPFLAELOQGLHODEFRRLDKGLFVLTDCGLNDTEL
KSAKDLVIDPEPHGSEAEKRAVTSVPMTMGFTGLSESAOTINMGSYDSMYS
MGTDNLPEPSGDFERITQYDRQYTSRAVAKEVLRATGTBPBGCAFLDCEILR
FRYPQVPEHSAEOPLEMAHPHLDLSWVTLIQTPCANGFVSLQAEVGAFTDIPYR
PDVALVFCGAIATLVTGQVKAPRHVAAPRRDQJAGSSRTSSVFELRPNDFTFSVP
LARECGFVSLDGEFATFQDMTIGNYNIRTSKA"
ORIGIN
Alignment Scores:
Pred. No.: 2,97e-136 Length: 936
Score: 1634.00 Matches: 310
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.68% Mismatches: 0
Query Match: 99.88% Indels: 0
DB: 6 Gaps: 0
10719236-1_232-1164 (1-311) x AX300747 (1-936)
Qy 1 MetAspThrThrValProThrPheSerLeuAlaGluLeuGlnGlyLeuHisGlnAsp 20
Db 1 ATGGACACGACGCTGCCACCTTCAGCTCGGCCGAATCTCCAGCAGGCGCTGCACCAAGAC 60
Qy 21 GluPheArgArgCysLeuArgAspArgGlyLeuPheTyLeuThrAspCysGlyLeuThr 40
Db 61 GAGTTCGCCAGGCTGCTCGAGGACAAAGGCGCTCTTCTATCTGACGAGACTGCGGCTGACC 120
Qy 41 AspThrGluLeuLeuSerAlaIlyAspIleValIleAspPhePheGlnHisGlySerGlu 60
Db 121 GACACCGAGCTGAATTCGGCCAAAGACCTCGTCATCTGACTTCTTGAGACGCGGCGAG 180
Qy 61 AlaGluTyArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeu 80

Db 181 GCGAGAAAGCGCGCGTCACTCCGCGCTCCCAACCAATGCGCGCGGCTTACCGGCGTG 240
Qy 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTySerAspTySerMetCys 100
Db 241 GAGTCGGAGACACCGCCCAATCAACCAATACCGGACCTACTCCGACTCGATGTCG 300
Qy 101 TySerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGlnHisGlyThr 120
Db 301 TACTGATGGGACCGCGGACCACTCTCCGTCGGTGAATCTGACCGGATTCGAGAC 360
Qy 121 GlnTyPheAspArgGlnTyThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
Db 361 CAGTACTTCGACCGCGCACTACCGCCCTCCGCGCGGTCCGCGGAGTCTCGGGCG 420
Qy 141 ThrGlyThrGluProAspGlyValValGlnAlaPheLeuAspCysGluProLeuArg 160
Db 421 ACCGGGACCGAGCCGAGCGCGGAGTCGAGGCTCTTCGACTCGACCCGCTGCGG 480
Qy 161 PheArgTyPheProGlnValProGlnHisArgSerAlaGluGluGlnProLeuArgMet 180
Db 481 TTCGCTACTTCGCGAGGTCCCGAGACCCGAGCGCGGAGAGCAGCCCTGCGAGTG 540
Qy 181 AlaProHisTyArgAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGly 200
Db 541 GCGCGGACCTACGACCTGTGATGCTCATTCCAGACAGACACCTGCGCGCAAGCG 600
Qy 201 PheValSerLeuGlnAlaGluValGlyValAlaPheThrAspLeuProTyArgProAsp 220
Db 601 TTCGCAAGCTTCGAGCGCGGAGGTGCGCGCGCTTCACGAGCTCCCTACCGTCCGAC 660
Qy 221 AlaValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGluValIlySala 240
Db 661 GCGTCTCTGTTCTTCTGCGCGCATCGCACTTGATCGGCGGCGGCAAGGCG 720
Qy 241 ProArgHisIlySalaIleAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260
Db 721 CCGCGGACCATGTGCGGCGCCCGCGGAGGACCAAGATGAGGAGCGGACGACCTCC 780
Qy 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280
Db 781 AGTGTGTTCTTCTCCGCTCCCAACGCGGACTTCGCTCCGCTCCGCTGCGCGCGAG 840
Qy 281 CysGlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrpIleGlyGly 300
Db 841 TCGGCTTCGATGTCAAGCTGAGCGGAGCGGACCGCGCTTCAGGATTGATCGGAGGC 900
Qy 301 AsnTyValAsnIleArgArgThrSerIlySala 311
Db 901 AACTACGTGAACATCCGCGGACATCCAGGCA 933
RESULT 4
STWCFDA 2494 bp DNA linear BCT 26-APR-1993
LOCUS S.clavuligerus isopenicillin N epimerase (cefi) gene and 2494
DEFINITION deacetoxycephalosporin C synthetase (DMOCs) gene, complete cds.
ACCESSION M32324.1 GI:153203
VERSION M32324.1 M24140
KEYWORDS
SOURCE Streptomyces clavuligerus
ORGANISM Streptomyces clavuligerus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE
1 Kovacevic, S., Weigel, B. J., Tobin, M. B., Ingolia, T. D. and Miller, J. R.
Cloning, characterization, and expression in Escherichia coli of
the Streptomyces clavuligerus gene encoding deacetoxycephalosporin
C synthetase
J. Bacteriol. 171 (2), 754-760 (1989)
MEDLINE 89123150
PUBMED 2644235
REFERENCE
2 (bases 1 to 1706)
Kovacevic, S., Tobin, M. B. and Miller, J. R.

TITLE The beta-lactam biosynthesis genes for isopenicillin N epimerase and deacetoxycephalosporin C synthetase are expressed from a single transcript in *Streptomyces clavuligerus*

JOURNAL J. Bacteriol. 172 (7), 3952-3958 (1990)

MEDLINE 90299822

PUBMED 16945525

COMMENT Original source text: S.clavuligerus DNA, clone POW80. Draft entry and computer-readable sequence for (1) kindly submitted by J.R.Miller, 26-FEB-1990, for release after publication. Location/Qualifiers

FEATURES

source 1..2494

mol_type="genomic DNA"

/db_xref="taxon:1901"

152..21706

/product="cefD mRNA"

284..1480

/note="isopenicillin N epimerase (cefD)"

/codon_start=1

/transl_table=1

/protein_id="AA26714.1"

/db_xref="GI:153204"

/translation="MAVADWEARGRMILDPVTNLTNGSGPLPRSAFERVTGFRALIAEPMPFLREVPALIMQARESLRILIGDPLRLATNVTAVNIVASRLAEGEILISDEYTPMRKQWVRVARRHGLRLRLPELSDPAEITAAVAASRRRLRFPSHYVSTTGLILPAALICEEARARITTVVDGAAAPFDLDLRIICDPYASGHKMLAPTVGFPHLAPRLBELSPVTSWAPYEPGSGPPAARDRGTGPGRLKRCGRDIDCPLATPESIDFOAELGPAIRARRRLTDHARRLLADRPGRITLLTPDSPLSGMVAAYRLPGTDAAELRGLMERFRLEAAVAEQPPVLRISANFYTEEBIDRLADALDALTGE"

1559..2494

/note="deacetoxycephalosporin C synthetase (DAOCS)"

/codon_start=1

/transl_table=1

/protein_id="AA26715.1"

/db_xref="GI:153205"

/translation="MDTTPPTFSLAELQQLHODEPFRCLRDKGLFYLTQGLTDTLSKARDVLDPEFGSEAKRAVTSVPVPMRSGFTGLSESTAOITNTGYSIDSMCVSMGTADNLFPSSDPERITQYDFDROYTSRAVARAVLATGEPGCGIAPFLDCEPLRFRTPOVPEHRSAEQPLRKMAHYDLSVNTLILQTPCANGVSVLSQAVGGVGLPLPRPDVAIVFCGAIATLVTGQVAPRHVAAPRRDQIAGSSRTSSVFLPRPADFTFSVPLARECGPVSIDGETATFQDWIGGNVYNIIRTSKA"

ORIGIN

Alignment Scores:

Pred. No.: 9.26e-136 Length: 2494

Score: 1634.00 Matches: 310

Percent Similarity: 100.00% Conservative: 1

Best Local Similarity: 99.68% Mismatches: 0

Query Match: 99.88% Indels: 0

DB: 1 Gaps: 0

10719236-1_232-1164 (1-311) x STMCFDA (1-2494)

Qy 1 MetAspThrThrValProThrPheSerLeuAlaGluLeuGlnGlyLeuHisGlnAsp 20

Db 1559 ATGACACGACGCGTCCACCTTCAGCTCGGCCCAACTCCACGAGGCTTCACACGAGAC 1618

Qy 21 GluPheArgArgCysLeuArgAspLysGlyLeuPheTyrLeuThrAspCysGlyLeuThr 40

Db 1619 GAGTTCGCGAGGTGTCTGAGGGACAAAGGCGCTCTTCTATCTGACGGACTGCGGATCTGACC 1678

Qy 41 AsnThrGluLeuLysSerAlaLysAspIleValIleAspPhePheGlnHisGlySerGlu 60

Db 1679 GACACCGAGCTGAAGTCCGCAAGGAACTCGTCATCGACTTCTTCGACACGAGGACGACG 1738

Qy 61 AlaGluLysArgAlaValIleThrSerProValProThrMetArgArgGlyPheThrGlyLeu 80

Db 1739 GCGGAGAAAGCGCGCTCACTCGCCCGTCCCGACCAACGCGCGGCTTACACCGGGCTG 1798

Qy 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys 100

Db 1799 GAGTCGAGAGACACGCGCCCAATGACCAATACCGGCACTACTCGACTCATCGATGTGC 1858

Qy 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTrpThr 120

Db 1859 TACTCGATGGGACCGCGGACCAACTTCTCCCTCGGTACTTCCAGCGGATCTGGACC 1918

Qy 121 GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140

Db 1919 CAGTACTTCGACCGCCAGTACACCGCTCCCGCGCGGTGCGCGGAGAGTCTCGGGGCG 1978

Qy 141 ThrGlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArg 160

Db 1979 ACCGGGACCGACCGGACCGCGCGGGGTGAGAGCTTCTCTCACTGCGAGCGGCTGCGG 2038

Qy 161 PheArgTyrPheProGluValProGluHisArgSerAlaGluGluGlnProLeuArgMet 180

Db 2039 TTCGCTACTTCCGCGAGGTCCCGAGCACCGGACGAGCGCCAGGAGACGCTTGGGATG 2098

Qy 181 AlaProHisTyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGly 200

Db 2099 GCGCGGACCTACGACCTGTGATGTTCACCTCATCCAGCAGACACCTTGGCCAAAGCGC 2158

Qy 201 PheValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrArgProAsp 220

Db 2159 TTGCTCAGCTCCAGGCGGAGGTGCGGCGGCGCTTCACGGAACCTGCGCTACCGCGGAC 2218

Qy 221 AlaValLeuValPheCysGlyValIleAlaThrLeuValThrGlyGlyGlnValLysAla 240

Db 2219 GCGGCTCTGCTCTTCTGCGGCGCATCGACCTGTGACCGCGGCGGCTCAAGGCC 2278

Qy 241 ProArgHisHisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260

Db 2279 CCGCGGACCATGTGCGGCGCGCGCGCGGAGGACAGTACGCGGACGACCGCACCTCC 2338

Qy 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280

Db 2339 AGGTGTCTTCTTCTCCGTCCCAACGCGGACTTACCTTCCGTCCGCGCGGCGGAG 2398

Qy 281 CysGlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrpIleGlyGly 300

Db 2399 TGGCGCTTCATGTACCTGACCTGGACGCGGAGACCGCCACGTTCCAGATTTGGATCGGGGC 2458

Qy 301 AsnTyrValAsnIleArgArgThrSerLysAla 311

Db 2459 AACTAGGTGAACATCGCGCCACATCCAGGCA 2491

RESULT 5

AY318742

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

15466573

2 (bases 1 to 936)

Hsu,J.S., Yang,Y.B., Wei,C.L. and Tsai,Y.C.

Cloning the DAOCS gene from *Streptomyces ambofaciens*

Unpublished

3 (bases 1 to 936)

Hsu,J.S., Yang,Y.B., Wei,C.L. and Tsai,Y.C.

Direct Submission

Submitted (08-JUN-2003) National Yang-Ming University, Institute of Biochemistry, 115, Li-Nong St. Sec. 2, Shih-Pai, Taipei 112, ROC

Location/Qualifiers

10719236-1_232-1164 (1-311) x AY317908 (1-1141)

QY 1 MetAspThrThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAsp 20
Db ATGGACGACGAGCGGTCCGACGCTTCCAGCTTCCGAGCTTCCAGAGGGGCTCCACACGAG 97

QY 21 GluPheArgArgCysLeuArgAspIleGlyLeuPheTyrLeuThrAspCysGlyLeuThr 40
Db GAGTTCGGCGAGCTGCTGCGGAGCAGGTCCTTCTTCATCGACGACGACGCGGCTGTGC 157

QY 41 AspThrGluLeuLeuSerAlaIleAspIleValIleAspPheGluHisGlySerGlu 60
Db GAGCGCGCATGAAATCCGCGCAGACGCTGCGCATTTGACTTTCGACGACGCGCAG 217

QY 61 AlaGluIysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeu 80
Db GAGGAGAGCGGGCGCGCAGCTCCACGATTCGCGCATTCGCGGGGCTTCACGCGGCTTG 277

QY 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys 100
Db GAGTCGAGAGACCGCACAGATCACACCGCGGAGCTTACTCCGACTTCTCGATGTGC 337

QY 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTyrThr 120
Db TACTGATGGGGCTGGCGGACAGCTCTTCCCGGGGGACTTCGACGGGCTGTGACG 397

QY 121 GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
Db CACTATTTCCGCGCGGTATGATGACCTCACCCAGAGAGTGGCGCGGCTCTCGAAGCG 457

QY 141 ThrGlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArg 160
Db ACCGGAACCGAACCGGTGGCGCGCTCGAGTCTTCTGCTGACTGCGAACCCTGCTGAG 517

QY 161 PheArgTyrPheProGlnValProGlnHisArgSerAlaGluGluGlnProLeuArgMet 180
Db TTCGGTACTTCCCGGAGGTCCCGAGGACCGCAGCGCGGACGACGCCCTCGCGAGT 577

QY 181 AlaProHisTyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGly 200
Db GCGCGCACTACGACCTGTGATGTGACGGATCCAGCAGACCCGCTGCCGCGAAGCGG 637

QY 201 PheValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrArgProAsp 220
Db TTCGTCAGCTTCCAGCGCGAGATCGCGCGCTTCTGAGACTTCCCGGCCAAGCGGAC 697

QY 221 AlaValLeuValPheCysGlyValIleAlaThrLeuValThrGlyGlyGlnValIysAla 240
Db GCGGTCTCTGCTTCTGCGGAGCCATCGGACCTGTGTGACCGGCGCGCAGGTGAAAGCC 757

QY 241 ProAlaGlnHisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260
Db CCGAGGACCATGATGCGCGCGCGCGGACCGGAGCCAGATCGCGGACGCGCGTACGTC 817

QY 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280
Db AGCGCTTCTTCTTCTTCCGCGCGCGCGCTTCACTTCTCGCTTCTTCTTCTTCTTCT 877

QY 281 CysGlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTyrIleGlyGly 300
Db TCGCGATTCGATATAGCTGTGACGCGGAGACCGCACGTTCAAGACGTCGATGAGGCG 937

QY 301 AsnTyrValAsnIleArgArgThrSerIysAla 311
Db AACTATGTGACATTCGACGAGCGTCCGAGCA 970

RESULT 7
LOCUS AY318743 936 bp DNA linear BCT 06-OCT-2004
DEFINITION Streptomyces chartreusis deacetoxycephalosporin C synthase gene.
ACCESSION AY318743

VERSION AY318743.1 GI:32492593
KEYWORDS Streptomyces chartreusis
SOURCE Streptomyces chartreusis
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
AUTHORS Heu, J.S., Yang, Y.B., Deng, C.H., Wei, C.L., Liaw, S.H. and Tsai, Y.C.
TITLE Family shuffling of expandase genes to enhance substrate specificity for penicillin G
JOURNAL Appl. Environ. Microbiol. 70 (10), 6257-6263 (2004)
PUBMED 15466573
REFERENCE
AUTHORS Heu, J.S., Yang, Y.B., Wei, C.L. and Tsai, Y.C.
TITLE Cloning the DAOCS gene from Streptomyces chartreusis
JOURNAL Unpublished
AUTHORS Heu, J.S., Yang, Y.B., Wei, C.L. and Tsai, Y.C.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-2003) National Yang-Ming University, Institute of Biochemistry, 115, Li-Nong St. Sec. 2, Shih-Pai, Taipei 112, ROC
FEATURES
source
1..936
/organism="Streptomyces chartreusis"
/mol_type="genomic DNA"
/db_xref="taxon:1969"
1..936
/note="DAOCS"
/codon_start=1
/transl_table=1
/product="deacetoxycephalosporin C synthase"
/protein_id="AAP85381.1"
/db_xref="GI:32492594"
/translation="MDTIVPVFSITIEQALHQAERPRCLAEKGVFVYNSGSEADL KSKDVAVDFFKGTGQKRAVTKADFTIRGTLGSSSTAKTNSGSTTDVSMKCS MGTSGNLEPTADEGVTGTFEDRYDASRVAQVLTSTTEPVGVEHVTCEPVR FRYEPGEHRSASEEPLRMAPHYDLSTVTLIQOTCPNPFVSLQVEDGFVDLPVR PDLVFCGAVATLATGALVKAPRHVAAAGROIYSSRTSSVFLRPADFTFSVP LAKRCGFDIGDGTATFQDWIMAGNYNLTAKTKA"
ORIGIN
Alignment Scores:
Pred. No.: 1,74e-106 Length: 936
Score: 1300.00 Matches: 238
Percent Similarity: 86.82% Conservative: 32
Best Local Similarity: 76.53% Mismatches: 41
Query Match: 79.46% Indels: 0
DB: 1 Gaps: 0
10719236-1_232-1164 (1-311) x AY318743 (1-936)

QY 1 MetAspThrThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAsp 20
Db ATGGATTCGACGAGGTCCCGTGTTCAGCTTCCAGAGCTTCCAGAGGCTCCACAGGCC 60

QY 21 GluPheArgArgCysLeuArgAspIleGlyLeuPheTyrLeuThrAspCysGlyLeuThr 40
Db GAGTTCGGCGCGCTGTGCGGAGATTCGCGGCTTCTTCACTGTGTGACGAGCGGCTGTGC 120

QY 41 AspThrGluLeuLeuSerAlaIleAspIleValIleAspPheGluHisGlySerGlu 60
Db GAGCGCGCATGAAATCCGCGCAGACGCTGCGCATTTGACTTTCGACGACGCGCAGC 180

QY 61 AlaGluIysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeu 80
Db GAGGAGAGCGGGCGCGCAGCTCCACGATTCGCGCATTCGCGGGGCTTCACGCGGCTTG 240

QY 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys 100
Db GAGTCGAGAGACCGCACAGATCACACCGCGGAGCTTACTCCGACTTCTCGATGTGC 300

QY 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTyrThr 120
Db TACTGATGGGGCTGGCGGACAGCTCTTCCCGGGGGACTTCGACGGGCTGTGACG 360

Db 301 TACTCCATGGGGACCTCCGGCAACCTGTTCCTCCCAACGGCGAGATTTCAGAACGGTATGAGAC 360
QY 121 GINTYrPhaAspArgGlnTYrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
Db 361 GGCTACTTCGATCGCATGTACAGACGCTCCCGGAGGTGGCCCGGAGGTCTCGAAGAGC 420
QY 141 ThrGlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArg 160
Db 421 ACCGGACCGAGCCGAGCCGGCTGTGAGCACTTTGTACCTTCGCAACCCGCTGCGG 480
QY 161 PheaTYrPheProGlnValProGluHisArgSerAlaGluGluGlnProLeuArgMet 180
Db 481 TTCGGTACTTCCTCCCGAGGTCCTCCGAGCACCGGACGGCGAGAGAGCCCTGCGAGT 540
QY 181 AlaProHisTYrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAangly 200
Db 541 GGGCGCATACGACTGTGACGGTACGAGCTTATCAACAGACCCGTGCCCCCAACGGC 600
QY 201 PheValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTYrArgProAsp 220
Db 601 TTCGTAGGCTTCAGAGTCCAGGTCCAGCGCGGCTTCGTGAGCTTCGCCGCGAGCGGAC 660
QY 221 AlaValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyValIysAla 240
Db 661 CTGTGCTGTCTTCTGCGGGGCGGTGCGGACCTTCGCAACCGGCGGCTGGTGAAGCA 720
QY 241 ProArgHisIleValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260
Db 721 CCACGGCACACGATCCCGCACCGGGCCGGAGTGTGGCGAGCCGTACGTCC 780
QY 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgIu 280
Db 781 AGCGTCTTCTCTCGCTCCCAACGGCACTTCACGTTCTCGTCCCTGCGCAAGCGG 840
QY 281 CysGlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrrIleGlyGly 300
Db 841 TCGGATTCGACATCGGCTGAGCGGCGACAGCGGACAGTTCAGAGACTGATCGGAGC 900
QY 301 AenTYrValAsnIleArgArgThrSerIysAla 311
Db 901 AACTACGTCACTTCGTACGAGAACCAAGCGG 933

RESULT 8
A42999 939 bp DNA linear PAT 06-MAR-1997
LOCUS A42999
DEFINITION Sequence 14 from Patent WO9504149.
ACCESSION A42999
VERSION A42999.1 GI:2298443
KEYWORDS
SOURCE
ORGANISM
Amycolatopsis lactamdurans
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Pseudonocardiaceae; Pseudonocardiaceae; Amycolatopsis.
REFERENCE
AUTHORS Bovenberg, R.A., Koekman, B.P., Hoekema, A., Van, D.L. and Verweij, J.
TITLE 1 (bases 1 to 939)
JOURNAL 3-(CARBOXYETHYLTHIO)PROPIONYL-7-ADCA
PATENT: WO 9504149-A 14 09-FEB-1995;
COMMENT Other publication PL 312747 960513
Other publication CA 2168004 950209.
FEATURES
location/Qualifiers
1..939
/organism="Amycolatopsis lactamdurans"
/mol_type="unassigned DNA"
/isolate="ATCC 27382"
/db_xref="taxon:1913"
1..939
/gene="CFBE"
1..>939
/gene="CFBE"
/codon_start=1
/transl_table=11

/protein_id="CAA02697.1"
/db_xref="GI:2298444"
/translation="MTDAITVFPDLAEIRBGLHSEFRHCLREKGVFLKGTGLAED
HASAEIADVDFDGTAEKAWTPITPTIRGVALGESESTAQITNGKTDYSMSY
SMGTADNLFPSAEPEKAWEDYFARMYRPSQVARYLVSGAEPEVGDALDCEPL
RLRYPEVEDRVAEEQPLRMAPHYDLSITVLHPTPCANGFSLQVNDGSDYVDI PA
OPGAVLVCAGVATVADGAIKAPENHVAAGADKRVSSRSVSFELRPNDGFRFSV
PRARCGPDVSIPIAETATFDWDIGNINIKTAAR"

ORIGIN

Alignment Scores:
Pred. No.: 1,84e-98 Length: 939
Score: 1210.00 Matches: 220
Percent Similarity: 83.87% Conservative: 40
Best Local Similarity: 70.97% Mismatches: 50
Query Match: 73.96% Indels: 0
DB: Gaps: 0

10719236-1_232-1164 (1-311) x A42999 (1-939)

QY 2 AspThrThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAspGlu 21
Db 7 GACGGACCGCTGCGACCTTCGATCTGCGCGAGCTGCGTGAAGGCTTCCACAGAGAG 66
QY 22 PheaArgCysLeuArgAspIysLeuPheTYrLeuThrAspCysGlyLeuThrAsp 41
Db 67 TTCGCGACCTGCTCGGAGAGAGGGCGTGTCTTACTTAAAGGACACCGGGCTCCCGAG 126
QY 42 ThrGluLeuYSerSerAlaYSerAlaValIleAspPhePheGlnHisGlySerGluAla 61
Db 127 GCGGACCAACGCTCGCGCGGAGATCGCGGTGAGCTTCTTCAGACCAAGGACCGAGGC 186
QY 62 GluIysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeuGlu 81
Db 187 GAGAAAGAGGGCGGTGATGATCCCGATCCCGACATCCGCGCGGCTGACCGCGGCTGAG 246
QY 82 SerGluSerThrAlaGlnIleThrAspThrGlySerTYrSerAspTYrSerMetCysTYr 101
Db 247 TCCGAGACACCGCCCAATCAAGACACCGCAAGTACACGACTTACTCGATGCTGTC 306
QY 102 SerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTrrThrGln 121
Db 307 TCGATGGGACCGCGGACCACTGTTCCCGACCGCGGCTTCGAGAAGCGGTGGAGGAC 366
QY 122 TyrPheAspArgGlnTYrThrAlaSerArgAlaValAlaArgGluValLeuArgAlaThr 141
Db 367 TACTTCGCGCGAGTGTACCGGCTTCGAGACGTCGCGCGAGTGTCTGACTGCGTTC 426
QY 142 GlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArgPhe 161
Db 427 GCGCGGAAACCGAGTCCGATGGAACCTTCTTCGACTGCGAAACCTTCGCTGCGCTG 486
QY 162 ArgTYrPheProGlnValProGluHisArgSerAlaGluGluGlnProLeuArgMetAla 181
Db 487 CGTACTTCCCGAGAGTCCCGAGAGATCGCGTGGCGGAGAGACCGCTCGGATGAGCC 546
QY 182 ProHisTYrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAanglyPhe 201
Db 547 CCGCATACGACCTTCGATGTCACCTGTATCCACCGAACCCCTTGGCGCAACCGGCTTC 606
QY 202 ValSerLeuGlnAlaGluValGlyValaPheThrAspLeuProTYrArgProAspAla 221
Db 607 GTCACTGCGAGTCCAGTGCAGCGGTCTATGTGACATCCGAGCCGAGCGGCGCG 666
QY 222 ValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyGlnValIysAlaPro 241
Db 667 GTGCTGCTGCTTCTGCGGGGCGGTGCGACGCTGTGGCGGACCGGCGGATCAAGCGGCC 726
QY 242 ArgHisIleValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSer 261
Db 727 AAGCACCACTGTGGCGCGCGCGCGGACCAAGCGGGTGGCGACAGCCGACCTCCAGC 786
QY 262 ValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCys 281

Db 787 GTGTTCTTCTCGCCGCCAAGCGGAGCTTCGCTTCGCTCGCGCGGAGCCAGGAGTGC 846
Qy 282 GlyPheaspValSerLeuaspGlyIuThrAlaThrPheGlnaspTrpIleGlyIAsn 301
Db 847 GGGTTCGACGTCAGATCCCGGCCGAGACCGGCACCTTCGACGACTGATCGCGGCAAC 906
Qy 302 TyrValasnIleArgArgThrSerIysAla 311
Db 907 TACATCAACATCCGGAAGACCGCCGCCGCC 936

RESULT 9
LOCUS A72870 939 bp DNA linear PAT 15-OCT-1999
DEFINITION Sequence 14 from Patent WO9504148.
ACCESSION A72870
VERSION A72870.1 GI:6063888
KEYWORDS
SOURCE Amycolatopsis lactandurans
ORGANISM Amycolatopsis lactandurans
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.
REFERENCE 1 (bases 1 to 939)
Bovenberg, R.A. and Koekman, B.P.
PROCESS FOR THE EFFICIENT PRODUCTION OF 7-ADCA VIA
2-(CARBOXYETHYLTHIO)ACETYL-7-ADCA AND
3-(CARBOXYMETHYLTHIO) PROPIONYL-7-ADCA
Patent: WO 9504148-A 14 09-FEB-1995;
JOURNAL GIST BROCADES NV (NL); BOVENBERG ROELOF ARY LANS (NL)
FEATURES
source
1..939
/organism="Amycolatopsis lactandurans"
/mol_type="unassigned DNA"
/isolate="ATCC 27382"
/db_xref="taxon:1913"
1..939
/gene="CEFE"
1..>939
/gene="CEFE"
/codon_start=1
/transl_table=11
/protein_id="CAB58463.1"
/db_xref="GI:6063888"
translation="MTDATVPTPDLAELREGHQEPHCLREKGVFLKGTGLAEAD
HASAREIADPDHGTAEAKKAVMPRTIRGRVAGLSESTAOITNGKXTDSMSY
SMGRADNLPPSAEFKAWEDYFARKYRASQVAKROYLTSVGAEPVGDALDSEPL
RLRYPEVDEBVAEOPLRMAPHYDLSIVLHQTPANGPVSIOVEVDGSYVDIPA
OPGAVLFCGAVATLVADGAIKAPGHVAAAGADKRVGSSRTSSVFLRPNGDFRSY
PRARCGDVSIAPETATFDWIGNYINIKTAAR"

ORIGIN
Alignment Scores:
Pred. No.: 1.84e-98 Length: 939
Score: 1210.00 Matches: 220
Percent Similarity: 83.87% Conservative: 40
Best Local Similarity: 70.97% Mismatches: 50
Query Match: 73.96% Indels: 0
Gaps: 0
DB: 6
10719236-1_232-1164 (1-311) x A72870 (1-939)

Qy 2 AspThrThrValProThrPheSerLeuAlaGluLeuGlnGlyLeuHiIeGlnaspGlu 21
Db 7 GACGGACCGCTGCCACCTTCATCTCGCCGAGCTCGCGAGGAGCTTGACACGAGAGAG 66
Qy 22 PheArgArgCysLeuArgAspIysGlyLeuPheTyrLeuThrAspCysGlyLeuThrAsp 41
Db 67 TTCGCCCACTGCTCGCGGAGAGAGGCGGTTCTACCTCAAGCGGACCGGGCTCGCGAG 126
Qy 42 ThrGluLeuLysSerAlaLysAspIleValIleAspPhePheGluHiIeGlySerGluAla 61
Db 127 GCGGACCAACGCTCGCGCGGAGATCGCGTGAATCTTTCGACACGAGCAGCGAGGCC 186

Qy 62 GluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeuGlu 81
Db 187 GAGAGAAAGGGGTGATGAGACCGCATCCGACCATCCGGCGGGGATACGCCGGCTGGAG 246
Qy 82 SerGluSerThrAlaGlnIleIerAsnThrGlySerTyrSerAspTyrSerMetCysTyr 101
Db 247 TCCGAGAGACACCGCGCAGATACGAAACACCGGCAAGTACACCACTACTCGATGTCGTAC 306
Qy 102 SerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluAlaGlyIleTyrThrGln 121
Db 307 TCGATGGGCACCCGGGCAACCTGTTCCGAGCGCATGTCAGAAAGGCGTGGAGAC 366
Qy 122 TyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAlaThr 141
Db 367 TACTTCCCGCGGAGTACCGCGCTTCGACAGAGCTCCGCGGACGCTGACCTCGGTC 426
Qy 142 GlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArgPhe 161
Db 427 GCGCGGAGACCCCGAGTCCGCGCATGACGCTTCCTCGACTGCGAACCCCTGCTGCCCTG 486
Qy 162 ArgTyrPheProGlnValProGluHiIeArgSerAlaGluGluGlnProLeuArgMetAla 181
Db 487 CGTACTTCCCGCGAGGTGCCGAGAGATCGGTGCGCGAGAGAGACCGCTGGAGTGGC 546
Qy 182 ProHisTyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGlyPhe 201
Db 547 CCGGACATACGACCTCTCGATGCTACCTGATCCACAGACCCCTTGCGGAAAGGGTTCC 606
Qy 202 ValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrArgProAspAla 221
Db 607 GTCAGCTGCGAGGTGCGAGGTGACAGGGTCTCATGTGACATCCCGCGGCGAGCGGGCGC 666
Qy 222 ValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyGluValAspAlaPro 241
Db 667 GTGCTGGTGTCTCGCGCGGTGGCGACGCTGTGTGCCACAGCGCGGATCAAGGCCGCC 726
Qy 242 ArgHisIleValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSer 261
Db 727 AAGCACACAGTGGCGCGCGCCCGCGGCAAGAGCGGTGGGACAGCCGACCTCCAGC 786
Qy 262 ValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCys 281
Db 787 GTGTTCTTCTCGCCGCCAAGCGGAGCTTCGCTTCGCTGCGCGCGGCGAGGAGTGC 846
Qy 282 GlyPheaspValSerLeuaspGlyIuThrAlaThrPheGlnaspTrpIleGlyIAsn 301
Db 847 GGGTTCGACGTCAGATCCCGGCCGAGACCGGCACCTTCGACGACTGATCGCGGCAAC 906
Qy 302 TyrValasnIleArgArgThrSerIysAla 311
Db 907 TACATCAACATCCGGAAGACCGCCGCCGCC 936

RESULT 10
LOCUS AR023763 939 bp DNA linear PAT 05-DEC-1998
DEFINITION Sequence 15 from patent US 5795733.
ACCESSION AR023763
VERSION AR023763.1 GI:3977057
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 939)
Bovenberg, R.Ary Lans., Koekman, B.Pieter., Hoekema, A., Van Der
Laan, J.Metke., Verweij, J. and De Vroom, E.
TITLE Process for the efficient production of 7-ADCA via
3-(carboxyethylthio) propionyl-7-ADCA
Patent: US 5795733-A 15 18-AUG-1998;
JOURNAL Location/Qualifiers
FEATURES
source
1..939
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 1,84e-98 Length: 939
Score: 1210.00 Matches: 220
Percent Similarity: 83.87% Conservative: 40
Best Local Similarity: 70.97% Mismatches: 50
Query Match: 73.96% Indels: 0
DB: 6 Gaps: 0

10719236-1_232-1164 (1-311) x AR023763 (1-939)

```
QY 2 AspThrThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHISGlnAspGlu 21
   7 GACGGACCGGTGGCGACCTTCGATCTGCGCGAGCTCGTAGGGCTTCGACCGAGAGAG 66
QY 22 PheArgArgCysLeuArgAspGlyGlyLeuPheTyrlleuThrAspCysGlyLeuThrAsp 41
   67 TTCGGCCACTGCTCGCGCAGAGAGGCGGTCTTACTCAAGGGCACCAGGCTCGCCGAG 126
QY 42 ThrGluLeuLysSerAlaLysAspIleValIleAspPheGluHISGlySerGluAla 61
   127 GCGGACCAAGCTCGCGCGGAGATCGCGGTGACTTCTTCGACCAAGGACCGAGGCC 186
QY 62 GluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeuGlu 81
   187 GAGAGAGAGCGGTGATGACCGCGATCCGACCATCGCGCGGTACCGCGGCTCGAG 246
QY 82 SerGluSerThrAlaGlnIleThrAsnThrGlySerTyrlSerAspTyrlSerMetCysTyrl 101
   247 TCCGAGACGACCGCGCAGATCAAGAACCGGCAAGTACACCACTACATGTCGATC 306
QY 102 SerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTyrlThrGln 121
   307 TCGATGGGACCGCGGACAACTGTTCCTCCAGCGCGAGTTCAGAGAGCGGTGGAGAGAC 366
QY 122 TyrPheAspArgGlnTyrlThrAlaSerArgAlaValAlaArgGluValLeuArgAlaThr 141
   367 TACTTGGCGGAGTGAACCGCGCTTCGAGAGAGTGGCGCGAGGTGTCGATCGGTC 426
QY 142 GlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArgPhe 161
   427 GCGCGGAGACCGCGAGGTGCGCATGAGACGCTTCCTGACTGCGAACCCCTGCTGCCCTGC 486
QY 162 ArgTyrlPheProGlnValProGluHISArgSerAlaGluGluGlnProLeuArgMetAla 181
   487 CGCTACTTCCCGAGGTGCGCGAGATCGCGGTGCGGAGAGAGCGCTGCGGATGAGC 546
QY 182 ProHISArgPheLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGlyPhe 201
   547 CCGGACTACGACTCTCGATCGTCAACCTGATCCACGACCCCTTGGCGGAGCGGATTC 606
QY 202 ValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrlArgProAspAla 221
   607 GTACAGCTCGAGGTGAGCGGATGAGCGGATCTGATGACATCCCGCGGACCGGCGGCG 666
QY 222 ValLeuValPheCysGlyValAlaIleAspThrLeuValThrGlyGlyGlnValLysAlaPro 241
   667 GTGCTGTGTCTTCGCGCGCGGTGCGAGCTGTGTGCGGACGCGCGCATCAAGGGCGCC 726
QY 242 ArgHISHisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSer 261
   727 AAGCAGCAAGTGGCGCGCGCGCGCGGAGCAGACAGCGGTGGCGAGCGCGACCTTCAGC 786
QY 262 ValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCys 281
   787 GTGTCTTCTCTCGCGCGCGCATCGGCTTCCTCGGTGCGCGGCGCGAGGTGC 846
QY 282 GlyPheAspValSerLeuAspGlyGlyValThrAlaThrPheGlnAspTrpIleGlyGlyAsn 301
   847 GGGTTCAGCTGACGATCCCGCGCGGAGACCGGACCTTCGAGCATGATCGCGGCGAAC 906
QY 302 TyrValAsnIleArgArgThrSerLysAla 311
   |||
```

DB 907 TACATCAACATCCGGAAGACCGCGCGCGCC 936

RESULT 11
LOCUS 191794
DEFINITION Sequence 14 from patent US 5726032.
ACCESSION 191794
VERSION 191794.1 GI:3936264
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 939)
AUTHORS Bovenberg, R.Ary, Ians, Koekman, B. Pieter, Hoekema, A., Van Der Laan, J. Meeske, Verweij, J. and De Vroom, E.
TITLE Process for the efficient production of 7-ADCA via 2-(carboxymethylthio)acetyl-7-ADCA and 3-(carboxymethylthio)propionyl-7-ADCA
JOURNAL Patent: US 5726032-A 14 10-MAR-1998;
FEATURES
source location/Qualifiers
1..939
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 1,84e-98 Length: 939
Score: 1210.00 Matches: 220
Percent Similarity: 83.87% Conservative: 40
Best Local Similarity: 70.97% Mismatches: 50
Query Match: 73.96% Indels: 0
DB: 6 Gaps: 0

10719236-1_232-1164 (1-311) x 191794 (1-939)

```
QY 2 AspThrThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHISGlnAspGlu 21
   7 GACGGACCGGTGGCGACCTTCGATCTGCGCGAGCTCGTAGGGCTTCGACCGAGAGAG 66
QY 22 PheArgArgCysLeuArgAspGlyGlyLeuPheTyrlleuThrAspCysGlyLeuThrAsp 41
   67 TTCGGCCACTGCTCGCGCAGAGAGGCGGTCTTACTCAAGGGCACCAGGCTCGCCGAG 126
QY 42 ThrGluLeuLysSerAlaLysAspIleValIleAspPheGluHISGlySerGluAla 61
   127 GCGGACCAAGCTCGCGCGGAGATCGCGGTGACTTCTTCGACCAAGGACCGAGGCC 186
QY 62 GluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeuGlu 81
   187 GAGAGAGAGCGGTGATGACCGCGATCCGACCATCGCGCGGTACCGCGGCTCGAG 246
QY 82 SerGluSerThrAlaGlnIleThrAsnThrGlySerTyrlSerAspTyrlSerMetCysTyrl 101
   247 TCCGAGACGACCGCGCAGATCAAGAACCGGCAAGTACACCACTACATGTCGATC 306
QY 102 SerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTyrlThrGln 121
   307 TCGATGGGACCGCGGACAACTGTTCCTCCAGCGCGAGTTCAGAGAGCGGTGGAGAGAC 366
QY 122 TyrPheAspArgGlnTyrlThrAlaSerArgAlaValAlaArgGluValLeuArgAlaThr 141
   367 TACTTGGCGGAGTGAACCGCGCTTCGAGAGAGTGGCGCGAGGTGTCGATCGGTC 426
QY 142 GlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArgPhe 161
   427 GCGCGGAGACCGCGAGGTGCGCATGAGACGCTTCCTGACTGCGAACCCCTGCTGCCCTGC 486
QY 162 ArgTyrlPheProGlnValProGluHISArgSerAlaGluGluGlnProLeuArgMetAla 181
   487 CGCTACTTCCCGAGGTGCGCGAGATCGCGGTGCGGAGAGCGCTGCGGATGAGC 546
QY 182 ProHISArgPheLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGlyPhe 201
   |||
```

Db 547 CCGCACTACGACCTCTCATCTGTCACCTCGATCCACGACCCCTTGCGGACGCGTTC 606
Qy 202 ValSerLeuGlnAlaGluValGlyAlaAphThrAspLeuProTyrArgProAspAla 221
Db 607 GTCAAGCTCTGACAGTGTGAGCGGTCTATGTGACATCCCGCGCACCCGCGCGCG 666
Qy 222 ValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlnValIleValAlaPro 241
Db 667 GTGCTGGTGTCTGCGCGCGGTGGAGCGCTGTGGCCGACGCGCGCATCAAGCGCGCC 726
Qy 242 ArgHisHisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSer 261
Db 727 AAGCCACCGATGTCGCGCGCGCGCGCGGACCAAGCGGAGGAGGAGCGACCTCCAGC 786
Qy 262 ValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCys 281
Db 787 GTGTTCTTCTGTCGCGCGCGCGCGCGCTTCTGCTGTCGCCGCGGCGCAGGAGTGC 846
Qy 282 GlyPheAspValSerLeuAspGlyGlnThrAlaThrPheGlnAspTyrIleGlyGlyAsn 301
Db 847 GGGTTCGACGTACGATCCGCGCGCGACCGCCACCTTCGACGACTGGATCGCGGAC 906
Qy 302 TyrValAsnIleArgArgThrSerIlyAsnAla 311
Db 907 TACATCAACATCCGGAAGACGCGCGCGCC 936
RESULT 12
A43001 942 bp DNA linear PAT 06-MAR-1997
LOCUS Definition Sequence 16 from Patent WO9504149.
ACCESSION A43001 GI:2298445
KEYWORDS
ORGANISM Amycolatopsis lactamdurans
SOURCE Amycolatopsis lactamdurans
Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales;
Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.
REFERENCE 1 (bases 1 to 942)
AUTHORS Bovenberg, R.A., Koekman, B.P., Hoekema, A., Van, D.L. and Verweij, J.
TITLE PROCESS FOR THE EFFICIENT PRODUCTION OF 7-ADCA VIA
JOURNAL 3-(CARBOXYETHYLTHIO)PROBIONYL-7-ADCA
Patent: WO 9504149-A 16 09-FEB-1995;
GIST BROCADES NV (NL)
COMMENT Other publication PL 312747 960513
Other publication CA 2168004 950209.
FEATURES
source
1..942
/organism="Amycolatopsis lactamdurans"
/mol_type="unassigned DNA"
/isolate="IC 411"
/db_xref="taxon:1913"
1..942
/gene="CEFE"
1..>942
/gene="CEFE"
/codon_start=1
/transl_table=11
/protein_id="CAA02698.1"
/db_xref="GI:2298446"
/translation="MTDAVPTPTDLAEIRGIIHQEFNRHCLREKGVLYAKGTGPAEA
DHASGREIADVPFDGTAEKKAVMTPITIRGYAGLESSTAOITWTKYTDVSM
YSMGADNLPSPAREKAMEDYFARMYASODVAROVLTSVGAPEVGMDFDCEPL
LRLRFPFEPEDRVABEOPLRMAFRYDLSITVLTHQTCAAGFVSLQYEVNYSYDIP
AOPGAVLFCGAVTLVADGAIKAPKHVAAPGADKRVGSSRTSSVFLRPVGDERS
VPRARECGDVSIPATETATFDWIGNYINIRKTYAAR"

DB: 6 Gaps: 1
10719236-1_232-1164 (1-311) x A43001 (1-942)
Qy 2 AspThrThrValProThrPheSerLeuAlaGluLeuGlnGlnIleuHisGlnAspGlu 21
Db 7 GAGCGACCGTGGCGACCTTCGATCTGCGCGAGCTGTGAGGAGGCTTGACACGAGAGAG 66
Qy 22 PheArgArgCysLeuAlaAspIlyGlyLeuPheTyrIleuThrAspCysGlyLeu---Thr 40
Db 67 TTCGCGACATGCTCGCGCGGAGAGGGGTGTTCATCTCAAGGCGACCGGAGTCCGCGC 126
Qy 41 AspThrGluLeuIlySerAlaIlyAspIleValIleAspPhePheGlnHisGlySerGlu 60
Db 127 GAGCGACACACGCTGCGCGCGGAGATGCGGTGATCTTTCACACACGCGACCGAG 166
Qy 61 AlaGluIlyAspAlaValThrSerProValProThrMetArgGlyPheThrGlyLeu 80
Db 187 GCCGAGAAAGAGGGGTGATGACGCCGATCCGACATCCGCGCGGCGGTACGCGGCGCTG 246
Qy 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys 100
Db 187 GCCGAGAAAGAGGGGTGATGACGCCGATCCGACATCCGCGCGGCGGTACGCGGCGCTG 246
Qy 247 GAGTCCAGAGCACCGCGCAGATCCAGACACCGCAAGTACACCGACTACTGATTCG 306
Db 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTyrThr 120
Qy 307 TACTCGATGGGACCGCGCGACACTTTCACGCGCGCGACAGTTTCAGAAAGCGGTGGAG 366
Db 121 GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
Qy 367 GACTACTTCCGCGGAGTACCGCGCTTCGACAGACGTCGCGCGCGAGTGCATCTCG 426
Db 141 ThrGlyThrGluProAspGlyGlyValGlnAlaPheLeuAspCysGlnProLeuLeuArg 160
Qy 427 GTCCGCGGAGACCGCGAGTGGATGAGCGCTTCTTCATCGACCGAACCTCTGCTGCGC 486
Db 161 PheArgTyrPheProGlnValProGluHisArgSerAlaGluGlnGlnProLeuArgMet 180
Qy 487 CTGGGCTACTTCCCCGAGAGTCCCGAGAGTCCCTGCGCGAGAGACCGCGCTGGATG 546
Db 181 AlaProHisTyrAspLeuSerMetValThrIleuIleGlnGlnThrProCysAlaAsnGly 200
Qy 547 GCCCGCACTACGACCTCTCGATGTCACCTGATCCACCAAGACCCCTTCGCGGAGCGG 606
Db 201 PheValSerLeuGlnAlaGluValGlyAlaAphThrAspLeuProTyrArgProAsp 220
Qy 607 TTCGTCAGCTGCGAGTGCAGGTGAGGTGACGGGTCTATGTGACATCCCGCGCACCGCGC 666
Db 221 AlaValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyGlnValIlyAla 240
Qy 667 GCGGTGCTGTGTCTCTGCGCGCGGTGGCGACGCTGTGGCGCACGCGGATCAAGGCG 726
Db 241 ProArgHisHisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260
Qy 727 CCCAAGCACACGATGCGCGCGCGCGCGCGGACCAAGCGGTGGCGACAGCGCACCTCC 786
Db 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280
Qy 787 AGCGTGTCTTCTGCGCGCGCGCGGAGACTTCGCTTCTGGTGGCGCGGCGCAAGGAG 846
Db 281 CysGlyPheAspValSerLeuAspGlyGlnThrAlaThrPheGlnAspTyrIleGlyGly 300
Qy 847 TGGCGGTTCGACGTCAGATCCGCGCGGAGACCGCGCATTCGACGACTGGATGCGGCGC 906
Db 301 AsnTyrValAsnIleArgArgThrSerIlyAsnAla 311
Db 907 AACTACATCAACATCCGGAAGACCGCGCGCC 936
RESULT 13
A72872 942 bp DNA linear PAT 15-OCT-1999
LOCUS Definition Sequence 16 from Patent WO9504148.
ACCESSION A72872

VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
A72872.1	GI:6063890		<i>Amycolatopsis lactamdurans</i> <i>Amycolatopsis lactamdurans</i> Bacteria: Actinobacteriales: Actinobacteridae: Actinomycetales Pseudonocardiales: Pseudonocardaceae; <i>Amycolatopsis</i> .	1 (bases 1 to 942)	Bovenberg, R. A. and Keekman, B. P.	PROCESS FOR THE EFFICIENT PRODUCTION OF 7-ADCA VIA 2-(CARBOXYMETHYLTHIO)ACETYL-7-ADCA AND 3-(CARBOXYMETHYLTHIO)PROPIONYL-7-ADCA Patent: WO 9504148-A, 16 09-FEB-1995;	GIST BROCADES NV (NL); BOVENBERG ROELOF ARIJ LANS (NL)	Location/Qualifiers 1..942

ORIGIN

Alignment Scores:

Pred. No.:	3.61e-97	Length:	29
Score:	1195.50	Matches:	191
Percent Similarity:	83.28%	Conservative:	400
Best Local Similarity:	70.42%	Mismatches:	51
Query Match:	73.07%	Indels:	1
DB:	6	Gaps:	1

10719236-1_232-1164 (1-311) X A72872 (1-942)

Qy	2	AspThrThrValProThrPheSerIleuAlaGlyLeuGlnGlnGlyLeuHisGlnAspGlu	21
Db	7	GAGCGACCGTGGCGACCTTTCGATCTTGGCGAGCTTCGTAAGAGCTTGCACAGAGAG	66
Qy	22	PheArgArgCysLeuAlaArgAspArgGlyLeuPheTyrLeuThrAspCysGlyLeu---Thr	40
Db	67	TTCGGCACACGCTGGCGAGAGAGCGCGTCTTCACTCAAGAGGACCGGGCTCCGGCC	126
Qy	41	AspThrGlyLeuLysSerAlaLysAspIleValIleAspPhePheGlnHisGlySerGlu	60
Db	127	GAGCGGACACACGCTCCGGCCCGGAGATCGCGTGGACTTCTTGACACAGGACCGAG	186
Qy	61	AlaGlnLysArgAlaValThrSerProValProThrMetAlaGlyGlyPheThrGlyLeu	80
Db	187	GCCGAGAGAGAGCGGTGATGAGCGCGACATCCGACCATCCGCGCGGGTAAACCGGGCTG	246
Qy	81	GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys	100
Db	247	GAGTCCGAGAGCACCGCGACATCAAGAACACCGGACAGTACACGACTTCCGATGTCG	306
Qy	101	TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTyrThr	120
Db	307	TACTCGATGGGACACCGGGACACCGTTCCTCCAGCGCGAGTTGAGAAAGCGGTGGAG	366
Qy	121	GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaValGluValLeuArgAla	140
Db	367	GACTACTTCCCGCGAGTATGACCGCGTTCGCGAGACGTGCGCGCGAGAGTCTACCTCG	426

QY	141	ThrglyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArg	160
Db	427	GTCCGCGGGGAACCCGAGGTGGCGATGACCGCTTCTCCGACTGGGAACCCCTGCGGCC	486
QY	161	PheArgTrpPheProGluValProGluHisArgSerAlaGluGluGluProLeuArgMet	180
Db	487	CTGCGCTACTCTCCCGAGGTGCGGAGGATGCGTGGCGGAGAGCGAGCCGCTGGGATG	546
QY	181	AlaProHisTrpAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGly	200
Db	547	GCCCGGCACTACGACTCTCGACTCTCAACCTGATCCACAGACCCCTTGGCCGAACGG	606
QY	201	PheValSerLeuGlnAlaGluValGlyValAlaPheThrAspLeuProTrpArgProAsp	220
Db	607	TTTCGTCACTGACAGGTGAGAGTGAAGGATCTATGTGACATCCCGGCGAGCGGGC	666
QY	221	AlaValIleuValPheCysGlyAlaAlaIleAlaThrIleuValThrGlyGlyGlnValAspAla	240
Db	667	GCGGAGCTGGGTGTTCTGGCGCGCGGTGGCGACGCTGGGCGGACCGCCGATCAAGGG	726
QY	241	ProArgHisHisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer	260
Db	727	CCCAAGCACACAGTGGCCGCGCGCGCGGACGAAGCGGGTGGGAGGACGCCGACCTCC	786
QY	261	SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu	280
Db	787	AGCGGTCTCTTCTGGCGGCCCAACGGGGAACTTCGCGCTTCTGGGTCCGCGGCGCAAGG	846
QY	281	CysGlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrpIleGlyGly	300
Db	847	TGCGGGTTCGACGTGACATCCCGGCGGAGACCGCACCTTCGACGACGTGATCGCGGCG	906
QY	301	AsnTrpValAsnIleArgArgThrSerIysAla	311
Db	907	AACACATCAACATCCGGAGAAGCGCCGCCCGC	939

RESULT 14

LOCUS	AR023762	942 bp	DNA	linear	PAT 05-DEC-1999
DEFINITION	Sequence 14 from patent US 5795733				
AK023762					

DEPOSITION	SEQUENCE	FROM	DATE	TIME
AR023762	AR023762.1	GI:3977056		

KEYWORDS
SOURCE
Unknown.

ORGANISM Unknown.
Unclassified.

REFERENCE
1 (bases 1 to 942)
Bovenberg, R. A. Y., Jans, J., Koekman, B., Pieter, J., Hoeckema, A., Van Der
1000 T. Matsche, W. J. and De Vries, P.

```

TITLE      Braden, C. M.; Bess, J. C. and de Vroom, E.
           Process for the efficient production of 7-ADCA via
           3-(carboxymethylthio) propionyl-7-ADCA
JOURNAL    Patent: US 5795733-A 14 18-AUG-1998;
FEATURES   Location/Qualifiers
           1..942
           /organism="unknown"
           /mol_type="unassigned DNA"

```

ORIGIN	Alignment Scores:	
Align. No.:	3,616-97	Length
Score:	1195.50	Matcher
Percent Similarity:	83.28%	Consent
Best Local Similarity:	70.42%	Mismatch
Query Match:	73.07%	Gaps:
DB:	6	Indels
10719235-1_232-1164 (1-311) x AR023762 (1-9421)		

Oy 2 AspharthaPProtrPhaserleuAglIteGlnGlnIleuHsIsgIlnaPglu 21
 Db 7 GACCCGACCGTgCCGACCTTCGATCTGgCCGACGTGCGTgAGGgCTTTCGACCCAGAGAGAG 66
 Oy 22 PheArGArGcysleuArGAsPluYsgIyleuPheTyrlleuThraAspCysglIleu---Thr 40

[illegible]

JOURNAL FEATURES			
source	Location/Qualifiers		
2- (carboxyethylthio) acetyl-7-ADCA and 3- (carboxymethylthio) propionyl-7-ADCA Patent: US 5726032-A 16 10-MAR-1998;			
1. .942	/organism="unknown"		
	/mol_type="unassigned DNA"		
ORIGIN			
Alignment Scores:			
Pred. No.:	3 61e-97	Length:	942
Score:	1195.50	Matche:	219
Percent Similarity:	83.28%	Conservative:	40
Best Local Similarity:	70.42%	Mismatches:	51
Query Match:	73.07%	Indels:	1
DB:	6	Gaps:	1
10719236-1_232-1164 (1-311) x 191795 (1-942)			
QY	2	ASPThrThxValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuH1EglnAspGlu	21
DB	7	GAGCGCAGCGGTGCGACCTTCGATCTGGCCGAGCTGGCTTGACACAGAGAG	66
QY	22	PheArgArgCysLeuArgAspGlyLeuPheTyrLeuThrAspCysGlyLeu---Thr	40
DB	67	TTCGCGCAGCTGCTGGCGAGAGAGCGGTGTCTTACCTCAAGGACCGGAGCTGCGGCC	126
QY	41	AspThrGluLeuLeuSerAlaLysAspLeuAlaLeuAspPhePheGlnH1EglnSerGlu	60
DB	127	GAGCGCAGCAGCAGCTTCGGCCGGAGATCGCGGTGACTTCTTGACACAGCAGCAG	186
QY	61	AlaGluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeu	80
DB	187	GCGGAGAGAGAGCGGTGATGACGCGCGATCCGACATCCGGCGGGGTACGCGGAGCTG	246
QY	81	GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys	100
DB	247	GAGTCCGAGAGCAGCCCGCAGATCAAGAACCGGCAAGTACACCGACTACTCGATGTCG	306
QY	101	TyrSerMetCylYThrAlaAspAsnLeuPheProSerGlyAspPheGluArgGlyLeuTrp	120
DB	307	TACTCGATGGACCCCGGACAACTGTTCCCGACGCCGAGTTTCGAAAGCGCTGGAG	366
QY	121	GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla	140
DB	367	GACTACTGCGCGGATGATACCGCGCTTCGACGAGCGTCCGGCGAGTGTGACTCG	426
QY	141	ThrGlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArg	160
DB	427	GTCGGCGCGGAGCCCGAGGTGCGATGAGACGGCTTCTCCGACTCGAACCCCTGCTCGC	486
QY	161	PheArgTyrPheProGlnValProGlnH1AspSerAlaGlnGluGlnProLeuArgMet	180
DB	487	CTGGCTACTTCCCGAGGTGCGCCGAGATGCGTGGCCGAGAGCAGCGCGCTCGGATG	546
QY	181	AlaProH1SerYAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGly	200
DB	547	GCCCCGACTAGCACTTCGATCGCACCGGATCCAGACCGACCCCTGGCGGAGAGGG	606
QY	201	PheValSerLeuGlnAlaGluValGlyValAlaPheThrAspLeuProTyrArgProAsp	220
DB	607	TTCGTGACCTCGCAGGTGAGGTGAGCGGTCTCTATGTGACATCCCGGCGAGCCGGC	666
QY	221	AlaValLeuValPheCysGlyAlaAlaLeuThrLeuValThrGlyGlnValAla	240
DB	667	GCGGTGCTGTGTTCGCGCGCGGGTGGAGCGCTGGTGGCCAGCGCGCATACAGCG	726
QY	241	ProArgH1SerYValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer	260
DB	727	CCCAAGACCACTGGCGCGCGCGCGGCGGAGCAAGCGGTGGGCAAGAGCGGACCTCC	786
QY	261	SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu	280

Db 787 AGCGTGTCTTCTCGCGCCCAACGGGGACTTCGCTCTCGGTGCCCGCGCCAGGAG 846
QY 281 CYESGlyPheAspValSerLeuAepGlyIuThrAlaThrPheGlnAspTTrpIleGlyGly 300
Db 847 TCGGGGTTCGACGTGACATCCCGGCGGAGACCGCACCCTTCGACGACTGGATCGGGC 906
QY 301 AsnTyValAsnIleArgArgThrSerTyAla 311
Db 907 AACTACATCAACATCCGGAAGACCGCGCGCC 939

Search completed: September 19, 2005, 16:57:53
Job time : 4929 secs

THIS PAGE BLANK (USPTO)

XX Yang Y, Wei C, Hsu J, Tsai Y;
XX WPI; 2003-78116/74.
DR P-PSDB; ADE47638.
XX
XX New mutated penicillin expandase having an expandase activity towards
PT penicillin G, which is at least at 2-fold higher than the activity of
PT wild-type expandase, useful for producing 7-aminodesacetoxycephalosporanic
PT acid (7-ADCA).
XX
XX
XX Disclosure; SEQ ID NO 1; 55pp; English.
XX
XX The invention relates to a novel mutated penicillin expandase having an
CC expandase activity towards penicillin G, which is at least at 2-fold
CC higher than the activity of wild-type expandase. The mutated penicillin
CC expandase is useful for producing 7-aminodesacetoxycephalosporanic acid.
CC The present sequence encodes the wild-type penicillin expandase.
XX
XX Sequence 1230 BP; 207 A; 455 C; 384 G; 184 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 6,43e-167 Length: 1230
Score: 1636.00 Matches: 311
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0
10719236-1_232-1164 (1-311) x ADE47637 (1-1230)
QY 1 MetAspThrThrValProThrPheSerLeuAlaGluLeuGlnGlyLeuHisGlnAsp 20
DB 232 ATGGACAGACGAGTCCCACTTCAGCCCTGGCCGAACTCCAGCAGGCGCTGACCAAGAC 291
QY 21 GluPheAcrGAcGAcLeuAcrGAcAspGlyGlyLeuPheTyrLeuThrAspGlyLeuThr 40
DB 292 GAGTTCGCCAGGTGCTCTGAGGAGCAAGGGCCCTCTTCTATCTGACGAGACTGGGGTGTGAC 351
QY 41 AspThrGluLeuLeuSerAlaLeuAsp11Leu11LeuAspPheGluHisGlySerGlu 60
DB 352 GACACCGAGCTGAACTCGCGCCAAAGACATCGTCATCGACTTCTTGAGCAGCGCGCCAG 411
QY 61 AlaGluValArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeu 80
DB 412 GGGGAAAGCGCGCGTCACTCGCCGCTCCACCAATGCGCGGCTTCCAGCGGCTG 471
QY 81 GluSerGluSerThrAlaGln1LeuThrAsnThrGlySerTyrSerAspTyrSerMetCys 100
DB 472 GAGTCGAGAGACACCGCCCAAGATCACTCAATCCGCAAGTACTCCGACTCGATGATGC 531
QY 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArg11LeuThr 120
DB 532 TACTCGATGGGACACCGCGCAACTCTTCCGTCGCGGTGACTTGGAGCGGATCTGAGAC 591
QY 121 GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
DB 592 CAGTACTTCGACCGCCAGTACACCGCTTCCCGCGGTGCGCGGAGGTCTTGGCGGCG 651
QY 141 ThrGlyThrGluProAspGlyGlyValGluAlaPheLeuAspGlyGluProLeuLeuArg 160
DB 652 ACCGGGACCGAGCCCGAGCGGGGTCCAGGCTTCTCTGACTGGAGCGCGTGTGGCG 711
QY 161 PheArgTyrPheProGlnValProGluHisArgSerAlaGluGlnGlnProLeuArgMet 180
DB 712 TTCCCTCTCTTCCCGCAGGTCCCGCAGGACCCGACGCGCGAGACAGCCCTGCGGATG 771
QY 181 AlaProHisTyrAspLeuSerMetValThrLeu11GlnGln1ThrProCysAlaAsnGly 200
DB 772 GCGCGGACTTACGACCTGTGATGGTCACTTCATCCAGACACCTTGGCGCAGCGAC 831
QY 201 PheValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrArgProAsp 220

DB 832 TTGCTGACCTCCAGCGGAGGTGCGGCGCGCTTCAAGGACCTGCCCTACCGTCCGAC 891
QY 221 AlaValLeuValPheCysGlyAla11Lea1ThrLeuValThrGlyGlnValValAla 240
DB 892 GCCCTCTCTCTTCTGCGCGCGCATCGACCTGTGTGACCGCGCGCTCAAGGCC 951
QY 241 ProArgHisHisValAla1ProArgArgAspGln1Lea1GlySerSerArgThrSer 260
DB 952 CCGCGGACCAATGTGCGGCCCCCGCAGGACCAATACCGGAGCAGCCGACCTCC 1011
QY 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280
DB 1012 AGTGTGTCTTCTCCGCTCCCAAGCGGACTTCACTTCTCCGTCGCGTGGCGCGGAG 1071
QY 281 CysGlyPheAspValSerLeuAspGlyGlnThrAlaThrPheGlnAspTyr11GlyGly 300
DB 1072 TGCCTCTTCATGTCACTCGACCGGAGAACCGCACGTTCCAGATTTGATCGGGGC 1131
QY 301 AsnTyrValAsn1LeaArgThrSerTyrAla 311
DB 1132 AACTACGTGAACATCGCGCGCACTCCAGGCA 1164
RESULT 2
AAS17241
ID AAS17241 standard; DNA; 936 BP.
XX
XX AAS17241;
AC
XX 12-MAR-2002 (first entry)
DT
XX
XX Streptomyces clavuligerus penicillin N expandase DNA sequence.
DE
XX
XX Penicillin N expandase; ring-expanding; penicillin G;
KM phenylacetate-7-ADCA; amino desacetoxyccephalosporanic acid; cephalosporin;
KM cephalalexin; penicillin V; ds.
OS
XX Streptomyces clavuligerus.
FX
FT Key Location/Qualifiers
FT CDS 1..936
FT /tag= a
FT /product= "Penicillin N expandase"
XX
XX WO200185951-A1.
XX
XX 15-NOV-2001.
XX
XX 09-MAY-2001; 2001WO-GB002047.
XX
XX 09-MAY-2000; 2000GB-00011185.
XX
XX (ACS-)- ACS DOBPAR UK LTD.
XX
XX Johnson RI, Newbert RW;
XX
XX WPI; 2002-075247/10.
XX
XX P-PSDB; AAU11044.
XX
XX New enzyme useful for ring-expanding penicillin G to produce phenylacetyl
PT -7-amino desacetoxyccephalosporanic acid, comprises penicillin expandase
PT having increased specificity for substrates such as penicillin G.
XX
XX
XX Disclosure; Page 36-37; 42pp; English.
XX
XX The present invention relates to new penicillin expandases modified to
CC improve the ring-expanding activity for a substrate which is not the
CC natural substrate of the unmodified expandase. The invention is useful
CC for ring-expansion of penicillin G to produce phenylacetate-7-ADCA (amino
CC desacetoxyccephalosporanic acid) and the phenylacetate side chain is
CC removed from phenylacetate-7-ADCA to produce 7-ADCA, which is useful as a
CC starting point for the production of a range of semi-synthetic
CC cephalosporins, most notably cephalalexin. The polynucleotide of the
CC invention is useful for producing a primer e.g. a PCR primer and as a

CC probe. The modification of the penicillin expandase enhances the activity
CC of the enzyme such as penicillin N expandase for penicillin G or V as a
CC substrate. The modified expandase has enhanced catalytic activity or
CC increased specificity for another substrate such as penicillin G. The
CC present nucleic acid sequence encodes the penicillin N expandase enzyme
CC of the invention

XX Sequence 936 BP; 154 A; 354 C; 284 G; 144 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	7,24e-167	Length:	936
Score:	1634.00	Matches:	310
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.68%	Mismatches:	0
Query Match:	99.88%	Indels:	0
DB:	6	Gaps:	0

10719236-1_232-1164 (1-311) x AAN92262 (1-936)

```
QY 1 MetAapThrThrValProThrPheSerLeuAlaGluLeuGlnGlyLeuHisGlnAsp 20
DB 1 ATGGAACACGACGAGTCCCACTTCAGCTGCGCCGAATCCAGACGAGCTGCACACGAGAC 60
QY 21 GluPheArgArgCysLeuArgAspLysGlyLeuPheTyrLeuThrAspCysGlyLeuThr 40
DB 61 GAGTCCGCGAGGTGCTGAGGAGCAAGGCGCTCTTCTATCTACGAGACTGCGGTCTGACC 120
QY 41 AapThrGluLeuLysSerAlaLysAspIleValIleAspPhePheGlnHisGlySerGlu 60
DB 121 GACACGAGCTGAAGTCCGACCAAGGACCTGTCATGACTTCTTGAGCAACGCGACGAG 180
QY 61 AlaGluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeu 80
DB 181 GCGGAGAACGCGCGCTCACTGCGCCGCTCCACCATCGCGCGCTTCACCGGAGTGG 240
QY 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys 100
DB 241 GAGTCCGAGACGACCGCCCAAGATCAATACCGGAGCTACTCCGACACTCGATGTC 300
QY 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGlnArgIleTyrThr 120
DB 301 TACTCGATGGGACCGCGGACCAACTCTTCCGTCGAGTGAATTCGAGCGGATCTGAGAC 360
QY 121 GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
DB 361 CAGTACTTGACCGCGACGACACCGGCTCCGCGGCTGCGCGGAGGCTCTGCGGAG 420
QY 141 ThrGlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArg 160
DB 421 ACCGGGACCGGACCGGACCGGCGGCTTCGAGCTGCGAGCGCGCTGCTGCGG 480
QY 161 PheArgTyrPheProGlnValProGlnHisArgSerAlaGluGlnProLeuArgMet 180
DB 481 TTCGGCTACTTCCCGGAGGTCCCGGAGCACCGGACCGGAGGAGAGAGCCCTGCGGAG 540
QY 181 AlaProHisTyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGly 200
DB 541 GCGCGGACTAGACCTGTCGATGTCACCTCATTCACAGACACCTCCGCGCAACGCG 600
QY 201 PheValSerLeuGlnAlaGluValGlyAlaAlaPheThrAspLeuProTyrArgProAsp 220
DB 601 TTCGTCAAGCTCCAGGCGCGAGGTCCGCGGCGGCTTCACGAGACTCGCCCTCCGAGAC 660
QY 221 AlaValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyValIleValAsp 240
DB 661 GCGGCTCTGCTCTTCTGCGGCGCATCGGACCTGTGACCGGCGGCGGAGTCAAGGCC 720
QY 241 ProArgHisIleValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260
DB 721 CCGCGGACCATGTCCGCGCGCGCGCGGACCAAGATGCGGCGGCGCGCACCTTC 780
QY 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280
```

```
DB 781 AGTGTCTTCTCCCTCCGTCACAGCGGACCTTCACTTCCTCGTCCGTCGCGCGGAG 840
QY 281 CysGlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTyrIleGlyGly 300
DB 841 TCGGCTCTGATGTCTGACGCTGAGCGGACCGGACCGGACGATTCAGATTGATCGGGGCG 900
QY 301 AsnTyrValAsnIleArgArgThrSerLysAla 311
DB 901 AACTACGTGAACATCCCGGACATCCAGGACA 933
```

RESULT 3

AAN92262
ID AAN92262 standard; DNA; 1230 BP.

AC AAN92262;

XX 25-MAR-2003 (revised)

DT 30-MAR-1990 (first entry)

DE Deacetoxycephalosporin C synthetase.

XX Deacetoxycephalosporin C synthetase; DAOCs; cephalosporin; penicillin N;
KW antibiotics.

XX Streptomyces clavuligerus.

XX Key Location/Qualifiers

FT mat_peptide 232..1167

FT /tag= a

XX /product= "DAOCs"

PN EP341892-A.

XX 15-NOV-1989.

PD 09-MAY-1988; 88US-00192273.

XX 09-MAY-1988; 88US-00192273.

XX (ELIT) LILLY & CO ELI.

PI Ingolia TD, Kovacevic S, Miller JR, Skatrud PL;

XX WPI: 1989-334231/46.

DR P-PSDB; AAP93215.

XX New recombinant DNA encoding de-acetoxy-cephalosporin C synthetase - for

PT increasing or inducing cephalosporin synthesis in microorganisms or in

PT vitro.

PS Claim 3; Page 27-28; 46pp; English.

XX The sequence encodes deacetoxycephalosporin C synthetase (DAOCs) which

CC catalyzes expansion of penicillin N. Cephalosporium strains transformed

CC with vectors carrying the gene will produce antibiotics more efficiently.

CC (Updated on 25-MAR-2003 to correct PW field.) (Updated on 25-MAR-2003 to

CC correct DR field.)

XX Sequence 1230 BP; 207 A; 455 C; 384 G; 184 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.87e-166	Length:	1230
Score:	1630.00	Matches:	310
Percent Similarity:	99.68%	Conservative:	0
Best Local Similarity:	99.68%	Mismatches:	1
Query Match:	99.63%	Indels:	0
DB:	1	Gaps:	0

10719236-1_232-1164 (1-311) x AAN92262 (1-1230)

```
QY 1 MetAapThrThrValProThrPheSerLeuAlaGluLeuGlnGlyLeuHisGlnAsp 20
DB 232 ATGGAACACGAGTCCCACTTCAGCTGCGCCGAATCCAGACGAGCTGCACACGAGAC 291
```

OY	21	GIupheAgaGyGleuAArgApLySgLyuePhEYrTleuThrAAspCySgLyueThr	40
Db	292	GAGTTCGCCAGGTGCTGAGGAGCAAGGACCTCTTCTATCTGACGAGCTGGCTTGACC	351
OY	41	AsPThrgLueLyuSerAlaLysAspLLeValLLeasPhePheGluHLeGlySerLlu	60
Db	352	GACACCGAGCTGAAATCGGCCAAGGACATCGTCACTTCCTTCGACACGCGACGAG	411
OY	61	AlaGluLySaRgAlaValThrSerProValProThMetAraGArgLyPheThrGlyLeu	80
Db	412	GCGGAGAAAGCGCGCGTCACTCGCCCGTCCCAACATGCGCGCGCTTCAACCGAGCTG	471
OY	81	GIuSerGIuSerThAlaGlnLeThraSerThrgLySerTYrSerAspTYrSerMetCys	100
Db	472	GAGTGGAGAGACCGCCAGATCACCAATACGGAGAGCTACTCCGACTACTCGAATGAGC	531
OY	101	TYrSerMetGlyThralaAspAsnLeuPheProSerGlyAspPheGluAraGlyLeTyrThr	120
Db	532	TACTCGAATGGGACCGCGGACAACTCTTCCGTCGGGAGCTTGAAGGATCTGGAC	591
OY	121	GlnTYrPheAspAraGlnTYrThralaSerAraGlaValAlaArgGluValLeuAraGla	140
Db	592	CAGTCTTCGACCGCGAGACACCCCTCCCGCGGATCGCCCGGAGAGTCTCTGGGGCG	651
OY	141	ThrgLYThrgLyuProAspGlyGlyValGluAlaPheLeuAspCySgLyProleuLeuAraG	160
Db	652	ACCGGAGACCGAACCGGACGGCGGGGTGAGAGCTTCTCTCGACTCGAGACCGCTGTGGG	711
OY	161	PheArgTYrPheProGlnValProGluHLeAsrSerAlaGluGluGlnProleuAraMet	180
Db	712	TTCGCGCTACTTCCCGAGGTCCCGGACACCGCAGCGCGAGAGACAGCCCTGGGAGATG	771
OY	181	AlaProHLeTYrAspLeuSerMetValThrLeuLLeGlnGlnThrProCyAlaAsnGly	200
Db	772	GCGCGGCACTAGACCTGTGCATGGTCACTTCATCACAGACACACCTCGCGCCAAAGAC	831
OY	201	PheValSerLeuGlnAlaGluValGlyGlyAlaPheThraAspLeuProTYrAraProAsp	220
Db	832	TTCGTCAAGCTTCCAGGCCAGAGTCCCGCGCGCTTCAAGACTCCCTACCGTCCGAG	891
OY	221	AlaValLeuValPheCySgLyAlaLeaLeaThrLeuValThrgLyGlyGlnValLysAla	240
Db	892	GCGGCTCTGTCTTCTTCGCGCGCCATCGGACCTCGTGACCGCGCGCAGGTCAAGGCC	951
OY	241	ProAraGHisHisValAlaAlaProAraArgAspGlnLeaGlySerSerAraThrgThSer	260
Db	952	CCCCGACCAACATGTGCGCGCCCCCGCGACGAGACAGATACGGGAGACAGCGCACTCC	1011
OY	261	SerValPhePheLeuAraGProAsnAlaAspPheThraPheSerValProleuAlaArgLyu	280
Db	1012	AGTGCTTCTTCTCTCGTCCCAACCGGACTTCACCTTCCGTCCGCTGGCGCGGAG	1071
OY	281	CysGlyPheAspValSerLeuAspGlyGluThralaThraPheGlnAspTYrTLeGlyGly	300
Db	1072	TGCGGCTTCGATGATCAGCTGTGACGCGAGACCGGACAGTTCAGAGATTGGATCGGGGCG	1131
OY	301	AsenTYrValAsnLLeAraGThrgThSerLysAla	311
Db	1132	AACTACGTGAAACATCGCGCACATTCAGGACA	1164
RESULT 4			
AAQ88789			
ID	AAQ88789 standard; DNA; 939 BP.		
XX	AAQ88789;		
AC			
XX	16-OCT-2003 (revised)		
DT	25-MAR-2003 (revised)		
DT	04-OCT-1995 (first entry)		
XX			
DE	Nocardia lactamdurans expandase gene (cefE).		
XX			

Primer: amplify; PCR; expandase gene; *cefe*; *Nocardia lactamdurans*; fungus;
Streptomyces clavuligerus; expression cassette; acyltransferase; fungus;
Penicillium chrysogenum; hybrid promoter; *Aspergillus nidulans*; 7-ADA;
 7-amino-desacetoxycephalosporanic acid; cephalosporin; antibiotic; ds.
 OS
 XX
 XX
 OS
 Amycolatopsis lactamdurans.
 PN
 WO9504148-A1.
 XX
 PD
 09-FEB-1995.
 XX
 PF
 29-JUL-1994; 94WO-EP002543.
 XX
 PR
 30-JUL-1993; 93EP-00202259.
 XX
 PR
 24-DEC-1993; 93EP-00203696.
 XX
 PA
 (KONN) GIST-BROCADES NV.
 PI
 Bovenberg RAL, Koekman BP, Hoekema A, Van Der Laan JM, Verweij J;
 XX
 XX
 DR
 WPI, 1995-082231/11.
 XX
 PT
 7-amino-desacetoxy-cephalosporanic acid prodn. in *Penicillium chrysogenum*
 by simultaneous expression of expandase and acyl-transferase.
 PS
 Example 1, Fig 8; 37pp; English.
 XX
 XX
 The nucleotide sequence of the *Nocardia lactamdurans* expandase gene *cefe*
 as amplified by primers A0484983-4. The sequence differs from the
 published sequence (Coque et al., Mol. Genet. 226 (1993), 453-458).
 CC
 The amplified sequence lacks 3 bases: a G from pos. 120, a C from pos.
 CC
 122 and a G from pos. 124 of the above published sequence. This causes a
 CC
 loss of the proline residue at amino acid pos. 41 of the corresponding
 CC
 amino acid sequence. The expandase gene (*cefe*) from either *Nocardia*
 CC
lactamdurans or *Streptomyces clavuligerus* were amplified by PCR and
 CC
 inserted into an expression cassette for simultaneous expression of the
 CC
cefe gene and the gene encoding an acyltransferase. The expression
 CC
 cassette is placed in the fungus *Penicillium chrysogenum*. Expression of
 CC
 the genes in the cassette is driven either by a *trp-lac* hybrid promoter
 CC
 or the promoter from the *Aspergillus nidulans* *gpdA* gene. The terminator
 CC
 is the 3'-end of the *P. chrysogenum* *pemE* gene (see A0484983-95). The
 CC
 cassette is used in the production of 7-amino-desacetoxycephalosporanic
 CC
 acid (7-ADA), an intermediate in the production of cephalosporin
 CC
 antibiotics. Note: the sequences shown in this patent are identical to
 CC
 those in patent WO 95/04149. (Updated on 25-MAR-2003 to correct PN
 CC
 field.) (Updated on 16-OCT-2003 to standardise OS field)
 XX
 SO
 Sequence 939 BP; 153 A; 325 C; 326 G; 135 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 5,57e-121 Length: 939
 Score: 1210.00 Matches: 220
 Percent Similarity: 83.87% Conservative: 40
 Best Local Similarity: 70.97% Mismatches: 50
 Query Match: 73.96% Gaps: 0
 DB: 2 Indels: 0
 10719236-1_232-1164 (1-311) x AA088789 (1-939)
 QY 2 AAPTTRTTTThValProThPhSeSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAspGlu 21
 Db 7 GACGGGACACCGTCCGACCTTCATCTGGCCGAGCTGCGTGAAGGCTTGCAACAGAGAGAG 66
 QY 22 PheATGATGCTyLeuAArgAspLysGlyLeuPheTYrLeuThrAspCYsGlyLeuThrAsp 41
 Db 67 TTCCCGCACTGCTGCGCGAGAGAGGGCGGTCTTCAACTCAAGGGCAACGGGGCTCGCCGAG 126
 QY 42 ThrGluLeuLysSerAlaLysAspIleValIleAspPhePheGluHisGlySerGluAla 61
 Db 127 GCGACACACGCTCGGCGCGGAGATCGCGGTGAGACTTCTTGACACCAACGACCGAGGCC 186
 QY 62 GluYEAAGAlaValThSerProValProThMeArgArgGlyPheThrGlyLeuGlu 81

Db 187 GAGAGAGGCGGTGATGACGCCGATCCGACCATCCGGGCGGGTACGCCGGCTGGAG 246
QY 82 SerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCysTyr 101
Db 247 TCCGAGAGCACCAGGATGACGAACACCGGCACTACCACTACCTGAGTGTGAC 306
QY 102 SerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluIleTyrPThrGln 121
Db 307 TCGATGGGCAACCGGCAACCTGTTCCCGCGCCGAGTTCGAGAAAGCGGTGGAGAGAC 366
QY 122 TyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValIleuArgAlaThr 141
Db 367 TACTTCGCGCGGATGACCGCTTCCGAGACGTCGCGGGAGGTGCTGACCTCGATC 426
QY 142 GlyThrGluProAspGlyValIleGluAlaPheLeuAspCysGluProLeuLeuAlaPhe 161
Db 427 GGGCGGAAACCCGAGGTGCGCATGAGACGCTTCTCGACCTGCAACCCCTGCGCCCTG 486
QY 162 ArgTyrPheProGlnValProGluIleArgSerAlaGluGlnProLeuArgMetAla 181
Db 487 CGCTACTTCCCGAGGTGCGCCGAGATCGCGTGGCGGAGAGCAGCCGCTGCGGATGCC 546
QY 182 ProHisTyrAspLeuSerMetValThrLeuIleGlnIleThrProCysAlaAsnGlyPhe 201
Db 547 CGGCACTACGACCTCTCGATCGTCACCTCGATCCACCAACCCCTTGGCGGAACGGGTTTC 606
QY 202 ValSerLeuGlnAlaGluValGlyIleAlaPheThrAspLeuProTyrArgProAspAla 221
Db 607 GTCAGCGCTCAGGTCGAGGTGAGACGGGTCTATGTGACATCCCGCGAGCGGCGGCG 666
QY 222 ValIleuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyGlnValIleAspAlaPro 241
Db 667 GTGCTGTGTCTTCGCGCGCGGCGGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 726
QY 242 ArgHisHisValAlaAlaIleProArgArgAspGlnIleAlaGlySerSerArgThrSerSer 261
Db 727 AAGCACCACTGAGCGCGCCGCGCGGAGCAAGCGGTGGGAGCGGACCTTCACAC 786
QY 262 ValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCys 281
Db 787 GGTGTTCTTCTGCGCCCAACGGGAGCTTCGCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 846
QY 282 GlyPheAspValSerLeuAspGlyIleThrAlaThrPheGlnAspTrpIleGlyIleAsn 301
Db 847 GGGTTCGAGCTGAGCATCCCGCGCGGAGACCGCACCTTCGACGACTCGATCGGCGCAAC 906
QY 302 TyrValAsnIleArgArgThrSerIleValAla 311
Db 907 TACATCAACATCCGGAAGACCGCGCGCGC 936
RESULT 5
AA082721
ID AA082721 standard; DNA; 939 BP.
XX
AC AA082721;
XX
DT 16-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 04-OCT-1995 (first entry)
XX
DE N. lactamdurans cefE gene PCR 1 product.
XX
XX N. lactamdurans; P. chrysogenum; cefE gene; PCR 1 product;
XX expression cassette; 7-amino-desacetoxy-cephalosporanic acid;
XX expandase gene; cephalosporin antibiotics; ss.
XX
XX Amycolatopsis lactamdurans.
XX
XX WO9504149-A1.
XX
XX 09-FEB-1995.
XX
XX 29-JUL-1994; 94WO-EP002544.
XX
XX

XX
PR 30-JUL-1993; 93EP-00202260.
PR 24-DEC-1993; 93BP-00203695.
XX
PA (KONN) GIST-BROCADES NV.
XX
PI Bovenberg RAL, Koekman BP, Hoekema A, Van Der Laan JM, Verweij J;
XX WPI, 1995-082232/11.
DR
XX 7-amino-desacetoxy-cephalosporanic acid prodn. in penicillium chrysogenum
XX PT - transformed with expandase gene, using 3,3'-thiodi:propionic acid as
XX PT side chain precursor and deacylation of intermediate.
XX
XX Example 1; Fig 8; 37pp; English.
XX
XX AA082720 is the N. lactamdurans cefE gene, it is compared with AA082721
XX CC the PCR 1 generated N. lactamdurans cefE gene. P. chrysogenum expression
XX CC cassettes for the N. lactamdurans cefE (expandase) gene were constructed
XX CC using the PCR 1 product. The transformed P. chrysogenum can now be used
XX CC for 7-amino-desacetoxy-cephalosporanic acid prodn. an intermediate for
XX CC cephalosporin antibiotics. (Updated on 25-MAR-2003 to correct PN field.)
XX CC (Updated on 16-OCT-2003 to standardise OS field)
SQ Sequence 939 BP; 153 A; 325 C; 326 G; 135 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 5,57e-121 Length: 939
Score: 1210.00 Matches: 220
Percent Similarity: 83.87% Conservative: 40
Best Local Similarity: 70.97% Mismatches: 50
Query Match: 73.96% Indels: 0
Gaps: 0
DB:

10719236-1_232-1164 (1-311) x AA082721 (1-939)
QY 2 AspThrTrpValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAspGlu 21
Db 7 GACGGACCGTGGCCACCTTCGATGTGGCCGAGCTGCTGAGGGCTTCACACGAGAGAG 66
QY 22 PheArgArgCysLeuArgAspIleuAspPheTyrLeuThrAspCysGlyLeuThrAsp 41
Db 67 TTCGCGCACTGCTGCGCGGAGAGGCGGTTCATCTCACTCAAGGAGCGGGCTGCGCAG 126
QY 42 ThrGluLeuIleSerAlaIleAspIleValIleAspPheIleGlyIleSerGluAla 61
Db 127 GCGGACCAAGCGCTCGCGCGGAGATCGCGGTGCTTCGACCAAGGACCGAGGCC 186
QY 62 GluIleAspAlaValIleThrSerProValProThrMetArgArgGlyPheThrGlyLeuGlu 81
Db 187 GAGAGAGAGCGGTGATGACCGCCGATCCGACCACTCCGCGCGGTGACCGCGGCTGGAG 246
QY 82 SerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCysTyr 101
Db 247 TCCGAGAGCACCAGGATGACGAACACCGGCACTACCACTACCTGAGTGTGAC 306
QY 102 SerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluIleTyrPThrGln 121
Db 307 TCGATGGGCAACCGGCAACCTGTTCCCGCGCCGAGTTCGAGAAAGCGGTGGAGAGAC 366
QY 122 TyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValIleuArgAlaThr 141
Db 367 TACTTCGCGCGGATGACCGCTTCCGAGACGTCGCGGGAGGTGCTGACCTCGATC 426
QY 142 GlyThrGluProAspGlyValIleGluAlaPheLeuAspCysGluProLeuLeuAlaPhe 161
Db 427 GGGCGGAAACCCGAGGTGCGCATGAGACGCTTCTCGACCTGCAACCCCTGCGCCCTG 486
QY 162 ArgTyrPheProGlnValProGluIleArgSerAlaGluGlnProLeuArgMetAla 181
Db 487 CGCTACTTCCCGAGGTGCGCCGAGATCGCGTGGCGGAGAGCAGCCGCTGCGGATGCC 546
QY 182 ProHisTyrAspLeuSerMetValThrLeuIleGlnIleThrProCysAlaAsnGlyPhe 201
XX

```
Db 547 CCGACCTACGACCTTCGATCGTACCTGATCCACGAGACCCCTTGGCGGACGCGTTC 606
Qy 202 ValSerLeuGlnAlaGluValGlyValAlaPheThrAspLeuProTyrArgProAspAla 221
Db 607 GTCAAGCCCTGAGGTGCGAGGTGCGCTTATGTGGACATCCCGGCGACCGCGCGCG 666
Qy 222 ValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlnValValAlaPro 241
Db 667 GTGCTGTCTGTGGCGGCGGTGGCGAGCGTGGCGGAGCGGCGGCGATCAAGCGGCC 726
Qy 242 ArgHisHisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSer 261
Db 727 AAGACCAACGTGCGCGCGCGCGCGGACCAAGCGGAGAGAGCGGACCGACCTCCAGC 786
Qy 262 ValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCys 281
Db 787 GTGTTCTTCTGCGCGCCCAACGGGGAATTCCGCTTCTGCTGCCCGCGGCGCAGGAGTGC 846
Qy 282 GlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrpIleGlyGlyAsn 301
Db 847 GGGTTCGACGTGACATCCCGCGCGGACCGGACCTTCGACGACTGGATCGCGGCAC 906
Qy 302 TyrValAsnIleArgArgThrSerIysAla 311
Db 907 TACATCAATCCGGAAGACCGCGCGCGCG 936
RESULT 6
AA082720
ID AA082720 standard; DNA; 942 BP.
XX
AC AA082720;
DT 16-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 04-OCT-1995 (first entry)
XX
DE N. lactamdurans cefE gene.
XX
KM N. lactamdurans; P. chrysogenum; cefE gene; expression cassette;
KM 7-amino-desacetoxy-cephalosporanic acid; expandase gene;
KM cephalosporin antibiotics; ss.
XX
OS Amycolatopsis lactamdurans.
XX
PN MO9504149-A1.
XX
PD 09-FEB-1995.
XX
PF 29-JUL-1994; 94MO-EP002544.
XX
PR 30-JUL-1993; 93BP-00202260.
XX
PR 24-DEC-1993; 93BP-00203695.
XX
PA (KONN ) GIST-BROCADES NV.
XX
PI Bovenberg RAL, Koekman BP, Hoekema A, Van Der Laan JM, Verweij J;
DR WPI; 1995-082232/11.
XX
PT 7-amino-desacetoxy-cephalosporanic acid prodn. in Penicillium chrysogenum
PT - transformed with expandase gene, using 3,3'-thiodi:propionic acid as
PT side chain precursor and deacylation of intermediate.
XX
XX Example 1; Fig 8; 37pp; English.
XX
XX AA082720 is the N. lactamdurans cefE gene, it is compared with AA082721,
XX the PCR 1 generated N. lactamdurans cefE gene. P. chrysogenum expression
XX cassettes for the N. lactamdurans cefE (expandase) gene were constructed
XX using the PCR 1 product. The transformed P. chrysogenum can now be used
XX for 7-amino-desacetoxy-cephalosporanic acid prodn. an intermediate for
XX cephalosporin antibiotics. (Updated on 25-MAR-2003 to correct PN field.)
XX (Updated on 16-OCT-2003 to standardise OS field)
```

```
XX
SQ Sequence 942 BP; 153 A; 326 C; 328 G; 135 T; 0 U; 0 Other;
Alignment Scores:
Score: 7.66e-120 Length: 942
Percent Similarity: 1199.50 Matches: 220
Best Local Similarity: 83.60% Conservative: 40
Query Match: 70.74% Mismatches: 50
DB: 73.32% Indels: 1
Gaps: 1
10719236-1_232-1164 (1-311) x AA082720 (1-942)
Qy 2 AspThrThrValProThrPheSerLeuAlaGluGlnGlnGlyLeuHisGlnAspGlu 21
Db 7 GAGCGGACCGTCCGACCTTCGATCTGGCCGAACTGCTAGAGGCTTGCACCGAGAGAG 66
Qy 22 PheArgArgCysLeuArgAspLysGlyLeuPheTyrLeuThrAspCysGlyLeu---Thr 40
Db 67 TTCGGCACCTGCTGGCGCGAGAGGGGCTGTCTACTCAAGGGGCAACCGGCGTGGCGGC 126
Qy 41 AspThrGluLeuLysSerAlaLysAspIleValIleAspPheGlnHisGlySerGlu 60
Db 127 GAGCGGACCAACGCGCTCGCGCGGAGATCGCGTGGACTTCTTCACACGACGACCGAG 186
Qy 61 AlAGluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeu 80
Db 187 GCCGAGAAAGAGGGGTGATGACGCGCATCCACACATCCGCGCGGATACGCGGAGCTG 246
Qy 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys 100
Db 247 GAGTCCGAGAGCACCGCGGAGATCAGACACCGGCAAGTACACCACTACTCGATGTGC 306
Qy 101 TyrSerMetGlyThrAlaAspLeuPheProSerGlyAspPheGlnValGlyIleTrpThr 120
Db 307 TACTCGATGGGCAACCGGGAACAACCTGTTCCGAGCGCAAGTTCAGAAAGCGTGGAG 366
Qy 121 GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
Db 367 GACTACTTCCGCGCGAGTACCGCGCTTCCGAGAGCTCCGCGGACAGTGTGACTCG 426
Qy 141 ThrGlyThrGluProAspArgGlyValGluAlaPheLeuAspCysGluProLeuLeuArg 160
Db 427 GTGCGCGGGAACCGGAGTCCGAGATGAGCGCTTCTCGACATCGCAACCCCTGTGCGC 486
Qy 161 PheArgGlyThrProGlnValProGlnHisArgSerAlaGluGlnPheLeuAspMet 180
Db 487 CTGGCTACTTCCCGAGGTGCGCGGAGATCGGTGGCCGAGAGCAGCGCTGGATG 546
Qy 181 AlaProHisTyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGly 200
Db 547 GCCCGGACTACGACCTTCGATCGTACCTGATCCACAGACCCCTTGGGGAACGGG 606
Qy 201 PheValSerLeuGlnAlaGluValGlyValAlaPheThrAspLeuProTyrArgProAsp 220
Db 607 TTGCTCAGCCTGCGAGGTGCGAGGTGAGGCTTATGTGACATCCCGGCGGACCGGCG 666
Qy 221 AlaValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlnValValAla 240
Db 667 GCGGTGCTGTGTTCGCGCGCGGTGCGACGTGTGGCCGACGCGGATCAAGCGG 726
Qy 241 ProArgHisHisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260
Db 727 CCCAAGCAACAGTGGCGCGCGCGCGGAGACAGCGGAGTGGGACACCGCACCTCC 786
Qy 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280
Db 787 AGCGTGTCTTCTGCGCGCCCAACGGGGACTTCGCGTCTCGGTGCGCGGCGCAGGAG 846
Qy 281 CysGlyThrAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrpIleGlyGly 300
Db 847 TGGCGGTTTCGACGTCAACATCCCGCGCGGAGACCGCACTTTCAGACGACTGATCGCGG 906
```

Qy	301	AGNTYTVAlaenllEArgArgrhSrLySAla 311
Db	907	AACTACATCATCATCGGAAGACCGCGCGCC 939
RESULT 7		
AA08693		
ID	AA08693	standard, cDNA; 2672 BP.
AC	AA08693;	
XX		
DT	16-OCT-2003	(revised)
DT	25-AUG-1996	(first entry)
DE	Cephameycin biosynthetic enzyme gene cluster.	
XX		
XX	Cephameycin; cephalosporin; antibiotic;	
KW	C-7 hydroxycephem methyltransferase; 3'-methylcephem hydroxylase;	
KW	3'-hydroxymethylcephem O-carbamoyltransferase; ss.	
XX		
OS	Amycolatopsis lactamdurans; strain LC411.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..70
FT		/*tag= a
FT		/label= pcbc
FT		/note= "3', region of pcbc gene"
FT	RBS	73..80
FT		/*tag= b
FT	CDS	84..794
FT		/*tag= c
FT		/label= ORF7
FT		/product= "C-7 hydroxycephem methyltransferase"
FT		/note= "(Claim 8, page 40)"
FT	misc_difference	504..602
FT		/*tag= d
FT		/note= "translated sequence of bases 503-602 given in the
FT		specification do not correspond to the Genetic Code"
FT	RBS	785..789
FT		/*tag= e
FT		/note= "putative RBS for ORF7/ORF8 cotranslation"
FT	CDS	802..1668
FT		/*tag= f
FT		/label= ORF8
FT		/product= "coupling protein"
FT		/note= "(Claim 7, page 39-40)"
FT	RBS	1678..1685
FT		/*tag= g
FT	CDS	1692..2624
FT		/*tag= h
FT		/label= ORF9
FT		/product= "3'methylcephem-hydroxylase"
FT		/note= "celf gene (Claim 5, page 39)"
FT	stem_loop	2634..2667
FT		/*tag= i
FT		/note= "stem-loop may correspond to a transcription
FT		terminator"
XX		
XX	W09529253-A1.	
XX		
XX	02-NOV-1995.	
XX		
XX	17-APR-1995;	95WO-US004801.
XX		
XX	22-APR-1994;	94US-00233605.
XX		
XX	(MERI) MERCK & CO INC.	
XX		
XX	Coque JJR, Enguita FJ, Fuente JL, Llarena FJ, Liras P, Martin JF;	
XX	WPI, 1995-382998/49.	
DR	P-ESDB; AAR92151, AAR92152, AAR92153, AAR92154.	
XX		
XX	DNA encoding late enzymes involved in cephamycin biosynthesis - and	

PI	related proteins, used to transform cells for cephamycin prodn.
XX	
PS	Disclouseure; Fig 3A-C; 61pp; English.
CC	An internal sequence (AAT08693) of a 5.4 kb BamHI cDNA fragment of the
CC	Nocardia lactamdurans cephamycin C gene cluster was isolated from a phage
CC	clone using a cefE gene probe. It contains the 3' region of the pbcC
CC	gene, open reading frames (ORF7 and ORF8) that code for a 2-component
CC	system comprising C-7 hydroxylase methyltransferase (AAR92152) and a
CC	coupling protein (AAR92153), and the cefE gene (ORF9) encoding 3'-
CC	methylcephem hydroxylase (AAR92154). Another isolated cDNA (AAT08694)
CC	included the downstream cmcH gene (ORF10) coding for 3'-
CC	hydroxymethylcephem O-carbamoyltransferase (AAR92156). These sequences
CC	can be used for prodn. of recombinant enzymes useful for industrial-scale
CC	cephamycin synthesis. (Updated on 16-OCT-2003 to standardise OS field)
XX	
SQ	Sequence 2672 BP; 441 A; 952 C; 901 G; 378 T; 0 U; 0 Other;
	Alignment Scores:
Pred. No.:	1,2e-94 Length: 2672
Score: .	972.50 Matches: 187
Percent Similarity:	73.68% Conservative: 37
Best Local Similarity:	61.51% Mismatches: 79
Query Match:	59.44% Indels: 1
Dbl:	2 Gaps: 1
10719236-1_232-1164 (1-311) x AAT08693 (1-2672)	
OY	2 AsphrthrValProthrrPheSerLeuAlaGluInGlnGlyLeuHisGlnAspGlu 21
Dd	1698 GACAAGACGGTAGCCGGTCTTGAGATGCCAACCTGCAGCGGTCCGCCAGGACAG 175
OY	22 PheaArgArCySleuAdgApIySGIyleuHetYrLeuThrAspCyGIyleuThzAsp 41
Dd	1758 TTCGGCGAGTGGCCCCCG---CGCGGGGTCTTTACTCACTACCGGGTAGCGGCCACCAG 181
OY	42 ThrGluLeuLysSerAlaLysAspRileValIleaspPhepegluHisGlySerGluAla 61
Dd	1815 CGAGACCCACCGGGTGGCCACCGACCGCATGGACTTCCTTCGCCCAAGGACGGCCGAG 187
OY	62 GluLysArgAlaValThrSerProValProThrmelAtagaGglyPheThrGlyLeuGlu 81
Dd	1875 GAGAAGCACGGCGGTGACACCAAGATCCCGACATGCGCGCCGGGTACTCGCGCTGAG 193
OY	82 SerGluSerThrAlaGlnIleThrAspThrGlySerYrSerAspYrSerMetCysTrpYr 101
Dd	1935 GCGGAAGACACCGCCACAGTACCAACACCGGCACCTACACCGATTACTCCATGCTGTAC 199
OY	102 SermetGlyThrAlaAspAsnLeuPheProSerGlyAspPhegluArgIleTrpThrGln 121
Dd	1995 TCAGTAGGGCANCGCGCGGCAACTGTCTCCGTCCGAAGAGTTCCAGTGGTGTGAGCGGAC 205
OY	122 TyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaargGluValLeuArgAlaThr 141
Dd	2055 TACTTCGCACACCTGTATCCGCGCGCGGAGAGACCGGCGCCCTGTGCTACACGCGCG 211
OY	142 GlyThrGluProasrGlyGlyValGluAlaPheLeuaspCysGluProLeuLeuArgPhe 161
Dd	2115 GGACCTTAACAGCGCGGAGACCTTGACACCTTGCTGACTGCGACCCGatGTGTGCGCTG 217
OY	162 ArgTyPhePheProGlnValProGluHisArgSerAlaGluGluGlnProLeuArgMetAla 181
Dd	2175 CGGACTTCCCAGGAGGTCCCGGACACCGCGCGCCCGGAGTACAGGCCACGGCGAGTGGCC 223
OY	182 ProHisrYrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGlyPhe 201
Dd	2235 CCGCATTACACGACCTGTCAATCATCACTTCAATCCACAGACCCCGGTGGCCAAAGTTTT 229
OY	202 ValSerLeuGlnAlaGluValGlyValAlaPheThrAspLeuProTyArgArgProAspAla 221
Dd	2295 GTCAAGCTTCGAGGCGGAAGTGAACGAGTGAAGTGTGAACCTGCGGCACTGAGAGACCC 235
OY	222 ValLeuValPhecysGlyAlaIleAlaThrLeuValThrGlyGlyGlnValIlysAlaPro 241

```
|||||
Db 2355 GTGGTGTCTGTGGCGCGCATCGCGCCCTGTGCACCCGAGGCGCGGTCCCGCC 2414
Oy 242 ATGHSIAIValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSer 261
Db 2415 AACCAACCACTGTCTCCCGCGGACCGAGCATGTCTCAAGGCGAGCGACCGACCTCGAGC 2474
Oy 262 ValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCys 281
Db 2475 GTGTTCTTCTGCGCGCGCTGACCGATTTCATCTTCTGCTGCGCGCGAGGAGATAC 2534
Oy 282 G1YPheAspValSerLeuAspGlyG1uThrAlaThrPheGlnAspTrpIleG1yG1yAsn 301
Db 2555 GCGCTCGACGTAGCGCTGGACATGGAGAGCGACCTTCGGGCACTGGATCGGAGCAAC 2594
Oy 302 TyrValAsnIle 305
Db 2595 TACGTACAGATG 2606

RESULT 8
AAT00591
ID AAT00591 standard; cDNA, 972 BP.
XX
XX AAT00591;
XX AC
XX 16-OCT-2003 (revised)
DT 13-APR-1996 (first entry)
DE cefR (ORF9) encoding 3'-methylcephemhydroxylase.
KW cephamycin C biosynthesis; cephem-carbamoyltransferase; hydroxylase;
KW methyltransferase; cluster; cmch; cmcl; cmcl; cefR; ss.
XX
XX Amycolatopsis lactamdurans.
XX
XX Key Location/Qualifiers
XX CDS 1..924
XX FT /*tag= a
XX FT /*note= "3'-methylcephemhydroxylase"
XX FT 934..967
XX FT /*tag= b
XX FT /*note= "putative terminator"
XX
XX MO9529253-A1.
XX
XX 02-NOV-1995.
XX
XX 17-APR-1995; 95WO-US004801.
XX
XX 22-APR-1994; 94US-00233605.
XX
XX (MERI ) MERCK & CO INC.
XX
XX Coque JJR, Enguita FJ, Fuente JI, Llaarena FJ, Liras P, Martin JF,
XX
XX MPI; 1995-382998/49.
XX
XX P-PSDB; AAR84175.
XX
XX DNA encoding late enzymes involved in cephamycin biosynthesis - and
XX related proteins, used to transform cells for cephamycin prodn.
XX
XX Claim 5; Page 39; 61pp; English.
XX
XX Three genes (AAT00591-93) located in the cluster of cephamycin C
XX biosynthesis in N. lactamdurans which encode deacetoxycephalosporin C
XX hydroxylase and two other proteins, which introduce the methoxyl group at
XX C-7 have been isolated and sequenced. The sequence of one of the latter
XX proteins resembles both cholesterol hydroxylases and methyltransferases
XX of different origins acting on hydroxyl groups present in aromatic or
XX quinine-type compounds; both proteins are required for the hydroxylation
XX at C-7 and the transfer of the methyl group from 8-adenosylmethionine to
XX the 7-hydroxycephem intermediate. In addition, the isolation, nucleotide
XX sequence, and the characterisation of a gene (AAT00590) encoding a 3'-
```

```
CC hydroxymethylcephem O-carbamoyl- transferase, designated cmch has been
CC shown. (updated on 16-OCT-2003 to standardise OS field)
XX
XX SQ Sequence 972 BP, 165 A; 358 C; 314 G; 135 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1.02e-94 Length: 972
Score: 967.50 Matches: 186
Percent Similarity: 73.84% Conservative: 37
Best Local Similarity: 61.59% Mismatches: 78
Query Match: 59.14% Indels: 1
DB: 2 Gaps: 1
10719236-1_232-1164 (1-311) x AAT00591 (1-972)
Oy 4 ThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuIleGlnAspGluPheArg 23
Db 4 ACGGTACCGGTCTTCTGACGATGGCGGAATGGCGAGCGGTCTGGCGAGGAGTTCCGC 63
Oy 24 ArgCysLeuArgAspGlyLeuPheTyrLeuThrAspCysGlyLeuThrAspTrpGlu 43
Db 64 GAGTGGGCGCGC--CGGGGGTCTTCTACTACCGGATACGCGGTACGCGCCACGACAGAC 120
Oy 44 LeuysSerAlaYsaAspIleValIleAspPheGlnIleGlySerGluAlaGluYs 63
Db 121 CACCGGTGGCCACCGACACCGGAGATGACTTCTTGCCCAAGGACGCGCGAGAGAG 180
Oy 64 ArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeuGluSerGlu 83
Db 181 CAGCGCTGACCAAGAGTCCGACCATCGCGCGGGATCTCGCGGTGAGGCGGAA 240
Oy 84 SerThrAlaGlnIleThrAsnTrpGlySerTyrSerAspTyrSerMetCysTyrSerMet 103
Db 241 AGCACGCGCCAGGTACACACACCGGACCTTACCTACGACTTCTCATTTGTTCTGATG 300
Oy 104 G1YThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTrpTrpGlnTyrPhe 123
Db 301 GGATTCGGCGCAACCTGTTCCCGTGAAGAGTGGAGTGGTCTGACGACGACTCTTC 360
Oy 124 AspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAlaThrGlyThr 143
Db 361 GACAGCTGTGACCGCGCGCGGACGAGACCGCGCGTGTGTGACCGCGCGGACAC 420
Oy 144 GluProAspGlyGluValGluAlaPheLeuAspCysGluPheLeuAspArgPheArgTyr 163
Db 421 TAGACGCGCGAGACCTCGACACCTCTGACTGCGACCCGGTCTGCGCTGCGGTAC 480
Oy 164 PheProGlnValProGluIleArgSerAlaGluGluGlnProLeuArgMetAlaProHis 183
Db 481 TTCCGGAGGTCCCGAGACACCGCGCGCGGACGAGTACAGCCAGCCGAGTGGCCGAC 540
Oy 184 TyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaIleAsnGlyPheValSer 203
Db 541 TACGACCTGTCCATCACCTTCTATCCACACCCCGTGGCCCAAGGTTTGTGAGC 600
Oy 204 LeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrArgProAspAlaValLeu 223
Db 601 CTGCAAGCCGAATGAGACGCTGAGATGTGAGCTCGCGACGCTCGACGCGGTGTC 660
Oy 224 ValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyGlnAlaYsaAlaProArgHis 243
Db 661 GTGCTGTGCGCGGATCGCGCGCTGTGTCAACGAGGCGCGGTGCGCGCCCAACAC 720
Oy 244 HisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgTrpTrpSerValPhe 263
Db 721 CACGTGTCTTCCCGGACCGGACGATGCTCAAGGCGACGACGCTTGCACGCTGTC 780
Oy 264 PheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCysGlyPhe 283
Db 781 TTCTGCGCGCGTGCACCGATTCTCTGCTGCGCGCGACGCGAGGAGTACGGCTTC 840
Oy 284 AspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrpIleGlyGlyAsnTyrVal 303
```

DB	841	GACGTCAAGCTCGACATGAGAGAGCGACCTTCGGCGACTGATGGGACCACTACTGTC	900
QY	304	Asn16 305	
DB	901	ACGATG 906	
RESULT 9			
AAQ10190	ID	AAQ10190 standard; DNA; 23666 BP.	
AC	AAQ10190;		
DT	25-MAR-2003 (revised)		
DT	27-MAR-1991 (first entry)		
DE	Cephalosporin antibiotic biosynthetic genes.		
XX			
KW	cephalosporin; antibiotic; S-(L-alpha-aminoadipyl)-L-cysteinyl-D;-		
KW	valine synthetase; isopenicillin N synthetase; isopenicillin N epimerase;		
KW	deacetoxycephalosporin C synthetase; beta-lactamase;		
KW	deacetoxycephalosporin C hydroxylase; ss.		
OS	Lyso bacter lactamgenus.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	6..2819	
FT		/*cag= 1	
FT		/label= ORF 9	
FT		2887..4200	
FT	CDS	/*cag= h	
FT		/label= ORF 8	
FT		4211..5443	
FT	CDS	/*cag= g	
FT		/label= ORF 7	
FT		5524..16692	
FT	CDS	/*cag= a	
FT		/label= ORF 1	
FT		16761..17741	
FT	CDS	/*cag= b	
FT		/label= ORF 2	
FT		17802..18761	
FT		/*cag= c	
FT		/label= ORF 3	
FT		18798..19739	
FT	CDS	/*cag= d	
FT		/label= ORF 4	
FT		19802..21061	
FT		/*cag= e	
FT		/label= ORF 5	
FT	CDS	21186..22343	
FT		/*cag= f	
FT		/label= ORF 6	
XX			
FN	JF02291274-A.		
XX			
PD	03-DEC-1990.		
XX			
PE	10-JAN-1990;	90JP-00003762.	
XX			
PR	01-FEB-1989;	89JP-00024710.	
XX			
PA	(TAKE) TAKEDA CHEM IND LTD.		
XX			
DR	WPI: 1991-018854/03.		
DR	P-ESDB; AAR10145, AAR10688, AAR10690, AAR10691, AAR10692,		
DR	AAR10693, AAR10694, AAR10695.		
XX			
PT	Pregn. of cephalosporin series antibiotics - comprises culturing		
PT	transformant of microbe transformed by plasmid contg. new DNA fragment.		
XX			
PS	Claim 4; Fig 1; 67pp; Japanese.		
XX			
CC	A fragment of the sequence comprising at least one of the cephalosporin		

CC	biosynthetic enzymes listed in the KEYWORDS can be cloned in a plasmid
CC	and used to transform microbes, such as bacteria or yeast. Although DNA
CC	is preferably isolated from <i>B. lactumensis</i> , similar sequences could be
CC	obtained from other bacteria containing genes coding for biosynthesis of
CC	cephalosporin series antibiotics. See also A010191-2. (updated on 25-MAR
CC	-2003 to correct PA field.)
SQ	Sequence 23666 BP; 3952 A; 8522 C; 7603 G; 3589 T; 0 U; 0 Other;
Alignment Scores:	
Pred. No.: 2,99E-90	Length: 23666
Score: 944.00	Matches: 118
Percent Similarity: 73.11%	Conservative: 45
Best Local Similarity: 58.36%	Mismatches: 8
Query Match: 57.70%	Indels: 0
DB: 2	Gaps: 0
10719236-1_232-1164 (1-311) x AAQ010190 (1-23666)	
OY	2 AspThrThrValProThrPheSerLeuAlaGluLeuGlnGlyLeuHisglnaspGlu 21
Db	17808 GATTACAGGCATTCGAGTATTTCGACTCGACGAACCTCGACGCGCTCGCATTCG 17866
OY	22 PheArgArgCysIleuArgAspGlyGlyLeuPhePheTyrlLeuThrAspCysGlyLeuThrAsp 41
Db	17868 TTCGGCAAACTCCGTTCGAGCGCGCGGTCTTAAGTGCGGAGGACGATTGATCAA 17922
OY	42 ThrGluLeuLysSerAlaLysAspLileAlaAspPhePhedGluHisglSerGluAla 61
Db	17928 ACCGACGACGCCAAGGCGATGAGCGCGGCTCATGACCTTTGAGAAACGACCGCGGAA 17988
OY	62 GluLysArgAlaValAlaThrSerProValProThrMetArgArgGlyPheThrGlyLeuGlu 81
Db	17988 CAGAAGAACCGCTCAGAAACCTGACCAGCGGAGCGGAGTTTCCGACCTGGAG 18047
OY	82 SerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCysTyr 101
Db	18048 GCCAGAGCACCGCGCGGATCACCAAGGCGGGAATTAACCGATCATCTGATGCTCAT 18107
OY	102 SerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTrpThrGln 121
Db	18108 TCGATCGAGCTGACGAGCAACTGTTCCTCCGTCGCGCGCTTGAAACCACTGGACCGGC 18167
OY	122 TyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAlaThr 141
Db	18168 TATTTGACCGCTTTCACGTCTGCACCCAGAGATGCGCGTTCGTCGGCGCGGTC 18222
OY	142 GlyThrGluProAspGlyGlyValAlaAlaPheLeuAspCysGluProLeuLeuArgPhe 161
Db	18228 GACCCCGGCTCGACGAGACGCGTAGCGAGTTCTTCGACTGCGATCCGTGTCGCTTC 18287
OY	162 ArgTyrPheProGlnValProGlnHisArgSerAlaGluGlnProLeuArgMetAla 181
Db	18288 CGCTTTTTCCCGAAGAGCGCGGAGAACCGCTGCGCGGACACAGAGCGCGGCATGGGG 18347
OY	182 ProHisTyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaasnGlyPhe 201
Db	18348 CCGATTACACCTGTTCACATCACACACCATTCAGACACACCTCGCCCAACGGCTTC 18407
OY	202 ValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrArgProAspAla 221
Db	18408 GTACGCTCGACGATCGACGCTCGACGCGCGCTTACGTGATCTGCGCGCGGCTCG 18466
OY	222 ValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyAlaValLysAlaPro 241
Db	18468 ATGGTGGTGTTCGCGCGCGGAGTGGCGCGCTGTTCGGGGCGGCAAGTCAAGCGCGG 18527
OY	242 ArgHisIstiValAlaIaAProArgTArgAspGlnIleAlaGlySerSerArgThrSerSer 261
Db	18528 CGCCACCAAGTTCCTTCGCGCTGATGACGACGACGCGGTGGAGACGCGGACCTTCAGC 18587
OY	262 ValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCys 281

Db 18588 GTGTCCTTCCTCGGCGCGAAGCCGGAACCTTCGCGTTCCTGCGCGCTCGCGCGCGCGCAAC 18647
Qy 282 G1YPheASPValSerLeuaspG1yGluThrAlaThrPheGlnaPTpI1eG1yG1yAsn 301
Db 18648 GGACTCGATGTCGATTCACCGCGCACACCCGACCTTCGCGAGTGATCGCGCGCAAC 18707
Qy 302 TTTValaen11eArg 306
Db 18708 TACGTCAACCTGCGC 18722
RESULT 10
AAN81128
ID AAN81128 standard; DNA; 996 BP.
XX
XX AAN81128;
AC
XX 24-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 29-OCT-1990 (first entry)
XX Deacetoxycephalosporin-C-synthetase and deacetylcephalosporin-C-synthetase structural gene.
DE
XX Deacetoxycephalosporin-C-synthetase; Deacetylcephalosporin-C-synthetase;
KW cephalosporin; DS DNA.
XX
OS Acetemonium chrysogenum.
XX
XX EP281391-A.
PN
XX 07-SEP-1988.
PD
XX 03-MAR-1988; 88EP-00301845.
PE
XX 04-MAR-1987; 87US-00021836.
PR
XX (BL1L) L1L1L & CO ELI.
PA
XX
PI Ingolia TD, Queener SM, Samson SM, Skatrud PL;
DR WPI; 1988-251947/36.
XX
PT Recombinant DNA from Cephalosporium acetemonium - encoding
PT deacetylcephalosporin C synthetase - di:acetoxycephalosporin C
PT synthetase activity.
XX
XX Disclosure; Page ?; 82DP; English.
PS
XX DNA sequences encoding deacetoxycephalosporin-C-synthetase and
CC deacetylcephalosporin-C-synthetase are used to construct vectors for a
CC wide variety of hosts, eg Escherichia coli, Penicillium and
CC Cephalosporium. Penicillium transformants may be used for cephalosporin
CC production. See also AAF80534, AAN81125, AAN81126, AAN81127 (Updated on
CC 25-MAR-2003 to correct PA field.) (Updated on 24-OCT-2003 to standardise
CC OS field)
XX
SQ Sequence 996 BP; 176 A; 345 C; 310 G; 165 T; 0 U; 0 Other:
Alignment Scores:
Pred. No.: 5,05e-91 Length: 996
Score: 933.50 Matches: 178
Percent Similarity: 70.83% Conservative: 43
Best Local Similarity: 57.05% Mismatches: 90
Query Match: 57.06% Indels: 1
DB: 1 Gaps: 1
10719236-1_232-1164 (1-311) x AAN81128 (1-996)
Qy 1 MetAapThrThrValProThrPheSerLeuAlaGluLeuGlnG1yLeuH1sGlnasp 20
Db 1 ATGACTTCCAGACCGCTCTTTCGTCGACGACCTTCAGAGCGGCAAGGTCTCTCAACC 60
Qy 21 GlupheArgArgCysleuArgAspIysG1yLeuPheTyLeuThrAspCysG1yLeuThr 40

Db 61 GAGCTCGCCGAGGCGCGTCAACACCAAGGATATTCTTACTTACGACGAGCGCGCTGTGC 120
Qy 41 AspThrGluLeuIysSerAlaIysAspI1eVal11aAspPheGluH1sG1ySerGlu 60
Db 121 GACGACGACCAACCTTCGCGCGCTGAGACGTGCTTTCCTTTCAGAAAGGAAACGAG 180
Qy 61 AlaGluIysArgAlaValThrSerProValProThrMetArgArgG1yPheThrG1yLeu 80
Db 181 GAGAGAAAGGCGCGTCAACGCTTCGCGCGACCGTCAACGCGCGCGCGCTTCTGCGCTC 240
Qy 81 GluSerGluSerThrAlaGlnI1eThrAsnThrG1ySerTySerAspTySerMetCys 100
Db 241 GAGTGGAGAGACACCGCGCTTCGTCACCGAGCGGGAAGTACTCGACTACTGACGTGCG 300
Qy 101 TySerMetG1yThrAlaAspAsnLeuPheProSerG1yAspPheGluArgI1eTPthr 120
Db 301 TACTCCATGGGCAATCGCGCGCAACCTGTTCGCCAACCGGGGCTTCAGAGACGTCTGGCAG 360
Qy 121 G1nTyPheAspArgG1nTyThrAlaSerArgAlaVal11aArgGluVal11eAspAla 140
Db 361 GACTACTTCGACCGCATGTACGCGGACCAAGATGTGCGCGCGCGCTTCTCAACTCT 420
Qy 141 ThrG1yThrGluProAspG1y--G1yValGluAlaPheLeuAspCysG1uProLeuLeu 159
Db 421 GTGGGCGCGCGCGCTCGCGCGGAGACATGTATGATCGTTCGAGTGCATCCCTCCTC 480
Qy 160 ArgPheArgTyPheProGlnValProGluH1sArgSerAlaGluGluGlnProLeuArg 179
Db 481 CGCTTACGTATCTTCCCGAGAGCGGACCGGTCCCGAAGAGAAACCCCTCCGC 540
Qy 180 MetAlaProH1sTyPheAspLeuSerMetValThrLeuI1eGlnI1nThrProCysAlaAsn 199
Db 541 ATGGAGACCCACCTACACCTATGACCATCACTCTGTGACCAAGACGCTGCGCAAC 600
Qy 200 G1yPheValSerLeuGlnAlaGluValG1yG1yAlaPheThrAspLeuProTyArgPro 219
Db 601 GGCTTCGTAGCTGTGACGTGCGAGGTGAGAGAAATTCGTGACCTCCCGACGCTCCCC 660
Qy 220 AspAlaValLeuValPheCysG1yAlaI1eAlaThrLeuValThrG1yG1yGlnVal 239
Db 661 GCGCCCATGTGCTTCTTCGCGCGCGGCGGACCTTGCGCGCGGCGGCAAGTCAAG 720
Qy 240 AlaProArgH1sH1sVal11aAlaProArgArgAspGlnI1eAlaG1ySerSerArgThr 259
Db 721 GCGCCCAAGACCGGGTCAAGTCTCCCGGCGGACCAAGCGCTCGGACACCCGACG 780
Qy 260 SerSerVal11ePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArg 279
Db 781 TCGAGCGTCTTCTTCGCGCGGAGCCGACCTTCAGCTTCAACGTCAGAGTCAAG 840
Qy 280 GluCysG1yPheAspValSerLeuAspG1yGluThrAlaThrPheGlnaPTpI1eG1y 299
Db 841 GAGGGGGTTCACGTCGCCATCCGTCGAGCGGACGAGCTTCAGAGAGTGGCTTGGC 900
Qy 300 G1yAsnTyValaen11eArgArgThrSerTyAla 311
Db 901 GGGAACTATGTCAACATCGGAGGATTAACCGCGC 936
RESULT 11
AAF60258
ID AAF60258 standard; DNA; 996 BP.
XX
XX AAF60258;
AC
XX 11-SEP-2003 (revised)
DT 27-APR-2001 (first entry)
XX
XX Cephalosporium acetemonium DACS/DAOS coding sequence.
KW Cephalosporium acetemonium; DACS/DAOS; expandase/hydroxylase;
KW Cephalosporium deacetoxycephalosporin C; synthetase/hydroxylase;
KW cephalosporin C; penicillin production; antibiotic production; ds.


```

XX OS Acremonium chrysogenum.
XX
XX PN US6180361-B1.
XX
XX PD 30-JAN-2001.
XX
XX PF 12-DEC-1988; 88US-00283429.
XX
XX PR 04-MAR-1987; 87US-00021836.
XX
XX PA (ELIL ) LILLY & CO ELI.
XX
XX PI Ingolia TD, Queener SM, Samson SM, Skatrud PL.
XX
XX DR MPI, 2001-181736/18.
XX
XX DR P-PSDB; AAB69572.
XX
XX PT Novel recombinant DNA sequence that encodes Cephalosporium
XX deacetoxycephalosporin C synthetase/hydroxylase polypeptide useful for
XX PT construction of expression vectors for improving efficiency and yield of
XX fermentation.
XX
XX PS Claim 2; Col 59-60; 67pp; English.
XX
XX CC The present sequence encodes the DACS/DACCS (expandase/hydroxylase)
XX CC polypeptide of Cephalosporium acremonium. The invention provides a novel
XX CC recombinant DNA sequence that encodes a Cephalosporium
XX CC deacetoxycephalosporin C synthetase/hydroxylase polypeptide. The DNA
XX CC sequence is useful for production of cephalosporin in a Penicillium host
XX CC cell. It is useful for the construction of expression vectors for
XX CC improving the efficiency and yield of fermentation involving a wide
XX CC variety of penicillin and cephalosporin antibiotic-producing organism.
XX CC The expression vector is useful in constructing strains for use by the
XX CC pharmaceutical industry and to introduce cephalosporin synthesizing
XX CC activities into high-level penicillin producing Penicillium strain. The
XX CC DNA sequence is useful for screening genomic libraries of organisms that
XX CC produce cephalosporin C or similar compounds for the presence of the
XX CC gene. It is useful for preparing labelled probes that are used to find
XX CC expandase-encoding DNA sequences in lactam-producing species. (Updated on
XX CC 11-SEP-2003 to standardise OS field)
XX
XX SQ Sequence 996 BP; 176 A; 342 C; 313 G; 165 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 5.05e-91 Length: 996
Score: 933.50 Matches: 178
Percent Similarity: 70.83% Conservative: 43
Best Local Similarity: 57.05% Mismatches: 90
Query Match: 57.06% Indels: 1
DB: 4 Gaps: 1

10719236-1_232-1164 (1-311) x AAE60258 (1-996)
QY 1 MetAspTh-ThrValProThrPheSerLeuAlaGluLeuGlnGlnGluHisGlnAsp 20
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ATGACTTCCAAAGTCCCGCTTCTGCTCTCGAGACTCAAGAAGCGGCAAGTCTCCAC 60
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 21 GluPheArgArgCysLeuArgAspIysGlyLeuPheTyrLeuThrAspCysGlyLeuThr 40
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GAGCTCGCGGAGCGCTGACCAACCAAGGATATCTTACTTACGAGAGCGGCGTGGTC 120
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 41 AspThrGluLeuLeuSerLeuAlaIysAspIleValIleAspPheGluHisGlySerGlu 60
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 GACGACGACACCACTCGGCGCGGTGAGAGCTGGTGACTTTTTCAGAAACGAAAGCGAG 180
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 AlaGluIysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeu 80
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 GAGGAGAAAGAGCGCCCTGACGCTCGGCCACCGTAAGCGCCCGCGCTTCTGACCTTC 240
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys 100
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 GAGTGGAGAGACCCCGCTGTCACCGAGACGGGCAAGTACTCGGACTACTCGACGTGC 300

```

```

QY 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluAlaGlyLeuThr 120
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 TACTCCATGGGATGCGGGGCAACTGTCCTCCGAACCGGGGCTTCGAGAGACTCTGGCAG 360
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 GACTACTTTCGACCGGATGACGGCCGACCAAGAGATGTCGGCGCGCGCTTCTCAACTCT 420
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 141 ThrGlyThrGluProAspGly--GlyValGluAlaPheLeuAspCysGluProLeuLeu 159
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 GTGGCGCGCCCGCTCGCGGGGAGGACATTGATGACTTCGTGAGTGGCATCTCCCTCC 480
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 160 ArgPheArgTyrPheProGlnValProGlnHisArgSerAlaGlnGlnGluProLeuArg 179
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 CGCTTACGTAATCTCCCGAAGTCCGGAGGACCGCTGCGCGGAAGAACCTCTCCGC 540
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 180 MetAlaProHisTyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsn 199
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 ATGGGACCCCACTAGCACTTATGACCATGACGCTCGGACACGACAGACCTCGGCCAAC 600
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 200 GlyPheValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrArgPro 219
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 GGCTTCGTGAGCGCTGACGTGCGAGGTGAGCGAGATTCGTGACCTCCGACGCTCCCC 660
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 220 AspAlaValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlnValIys 239
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 GCGGCAAGTGTGTTTGTGGCGCGCGGTGCGCACCTCGGCGACGCGGCGGCAAGGTTCAG 720
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 240 AlaProArgHisIleValAlaAlaProArgArgArgGlnIleAlaGlySerSerArgThr 259
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 GCGGCCAAGACACCGGCTGAAGTCTCCGGGGCGACGACGCGCTGCGAGACCGCGACG 780
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 260 SerSerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArg 279
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 TCGAGCGCTTCTTCTCTCGCGCGGACCGACCTTCACCTTCACGTCACGACGTGCGAG 840
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 280 GlyCysGlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspThrIleGly 299
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 GAGTGGGTTTCACAGTTCGCAATCCCTCGGAGCGCACGAGTTCAGGAGTGGCTTGGC 900
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 300 GlyAsnTyrValAsnIleArgArgThrSerIysAla 311
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 GGAACATATGTCACATCGGAGGATTAAGCCGCGC 936
   1::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
AAN81125
ID AAN81125 standard; DNA; 1509 BP.
XX
XX AC AAN81125;
XX
XX DT 24-OCT-2003 (revised)
XX DT 25-MAR-2003 (revised)
XX DT 29-OCT-1990 (first entry)
XX
XX DE Deacetoxycephalosporin-C-synthetase and deacetylcephalosporin-C-
XX synthetase gene.
XX
XX KW Deacetoxycephalosporin-C-synthetase; Deacetylcephalosporin-C-synthetase;
XX cephalosporin; DS DNA.
XX
XX OS Acremonium chrysogenum.
XX
XX PN EP281391-A.
XX
XX PD 07-SEP-1988.
XX
XX PF 03-MAR-1986; 88EP-00301845.
XX
XX PR 04-MAR-1987; 87US-00021836.
XX
XX PA (ELIL ) LILLY & CO ELI.
XX

```

```
PI Ingolia TD, Queener SW, Samson SM, Skatrud PL;
XX WPI: 1988-251947/36.
DR P-PSDB; AAB80534.
XX
PT Recombinant DNA from Cephalosporium acremonium - encoding
PT deacetylcephalosporin C synthetase - di:acetoxy:cephalosporin C
PT synthetase activity.
XX
PS Disclosure; Page 7; 82pp; English.
XX
CC DNA sequences encoding deacetoxycephalosporin C-synthetase and
CC deacetylcephalosporin C-synthetase are used to construct vectors for a
CC wide variety of hosts, eg Escherichia coli, Penicillium and
CC Cephalosporium. Penicillium transformants may be used for cephalosporin
CC production. See also AAB80534, AAN81126, AAN81127, AAN81128 (updated on
CC 25-MAR-2003 to correct PA field.) (Updated on 24-OCT-2003 to standardise
CC OS field)
XX
SQ Sequence 1509 BP; 327 A; 456 C; 429 G; 297 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8,99e-91 Length: 1509
Score: 933.50 Matches: 178
Percent Similarity: 70.83% Conservative: 43
Best Local Similarity: 57.05% Mismatches: 90
Query Match: 57.06% Indels: 1
DB: 1 Gaps: 1

10719236-1_232-1164 (1-311) x AAN81125 (1-1509)

QY 1 MetAspThrThrValProThrPheSerLeuAlaGluGlnGlnGlyLeuHisGlnAsp 20
DB 400 ATGACTTCCAAAGTCCCGTCTTCTGCTCGACGACTTCAAGAGCGGCAAGGCTCTCACC 459
QY 21 GluPheArgArgCysLeuArgAspLysGlyLeuPheTyLeuThrAspCysGlyLeuThr 40
DB 460 GAGCTCGCGGCGCGCTCACACCAAGGATCTTCTTACTTACCGAAGAGCGGCTGTC 519
QY 41 AspThrGluLeuLysSerAlaLysAspIleValIleAspPheGluHisGlySerGlu 60
DB 520 GACGCGACCAACACCTCGCGCGGTGACGTCGCTTTCATTTTCAGAACGGAAGCGAG 579
QY 61 AlaGluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeu 80
DB 580 GAGGAGAAAGAGGCGCGTGAAGCTCGCCGCAACGCGCGCGGCTTCTCTCCCTC 639
QY 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTySerAspTySerMetCys 100
DB 640 GAGTGGAGAGACCGCGCGTGTCAACGAGCGGCAAGTACTCGGACTACGACGAGC 699
QY 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleThrThr 120
DB 700 TACTCCATGGGATCGCGCGCAACCTGTTCCGAACCGGCGGCTTCGAGGACCTCGGAG 759
QY 121 GlnTyrrPheAspArgGlnTyrrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
DB 760 GACTACTTGCAGCGGATGATGCGCGCAGCAAGATGTGCGCGCGCGCTTCAACTCT 819
QY 141 ThrGlyThrGluProAspGly--GlyValGluAlaPheLeuAspCysGluProLeuLeu 159
DB 820 GTGGGCGCGCGCTCGCGGAGAGACATGATGACTTGTGAGATGGGATCCCTCTTC 879
QY 160 ArgPheArgTyrrPheProGlnValProGluHisArgSerAlaGluGluGlnProLeuArg 179
DB 880 CGCCTACGGTACTTCCCGCAAGTGGCGGAGACCGCGTCCGCGAAGAGAACCTCCCGC 939
QY 180 MetAlaProHisLeuTyrrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsn 199
DB 940 ATGGGACCCCATGACCTATGACATCAACGATCGATGTCACCAAGACCTCCGCCAAC 999
QY 200 GlyPheValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrrArgPro 219
```

```
DB 1000 GGCCTTCGAGCCTGACGTGCGAGGTGAGCGAAGATTGCTGACCTCCGAGCGTCC 1059
QY 220 AspAlaValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlnValLys 239
DB 1060 GCGCCCATGTGCTGCTTCTCGCGCGCGGTGCGGACCTGCGCACGGCGCGCAAGTCAAG 1119
QY 240 AlaProArgHisLeuValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThr 259
DB 1120 GCGCCCAAGCACCGGATCAAGTCTCCGCGCGGACCAACGCGCTCGGACAGCCGACG 1179
QY 260 SerSerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArg 279
DB 1180 TCGAGCGCTTCTTCTCGCGCGCGAAGCCGACCTTCAAGTTCAGTTCAGTTCAGG 1239
QY 280 GluCysGlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrpIleGly 299
DB 1240 GAGGGGGTTTCAACGTCGCCATCCGTCGAGCGACGACGAGCGTTCAGAGAGTGGCTTGGC 1299
QY 300 GlyAsnTyrrValAsnIleArgArgThrSerLysAla 311
DB 1300 GGGAACTATGTCAATCGGAGGATTAACCGCGC 1335

RESULT 13
AAF58690
ID AAF58690 standard; DNA; 1511 BP.
XX
XX AAF58690;
AC
XX 11-SEP-2003 (revised)
DT 27-APR-2001 (first entry)
XX
DE Cephalosporium acremonium DACS/DACS gene.
XX
XX Cephalosporium acremonium; DACS/DACS; expandase/hydroxylase;
XX KM Cephalosporium deacetoxycephalosporin C; synthetase/hydroxylase;
XX KW cephalosporin C; penicillin production; antibiotic production; ds.
XX OS Acremonium chrysogenum.
XX
XX US6180361-B1.
XX
XX 30-JAN-2001.
XX
XX 12-DEC-1988; 88US-00283429.
XX
XX 04-MAR-1987; 87US-00021836.
XX
XX (EHL ) LILLY & CO ELI.
XX
XX Ingolia TD, Queener SW, Samson SM, Skatrud PL;
XX
XX WPI: 2001-181736/18.
XX P-PSDB; AAB69572.
XX
XX Novel recombinant DNA sequence that encodes Cephalosporium
XX PT deacetoxycephalosporin C synthetase/hydroxylase polypeptide useful for
XX PT construction of expression vectors for improving efficiency and yield of
XX PT fermentation.
XX
XX Disclosure; Col 7-12; 67pp; English.
XX
XX The present sequence is the DACS/DACS (expandase/hydroxylase) gene of
XX CC Cephalosporium acremonium. The invention provides a novel recombinant DNA
XX CC sequence that encodes a Cephalosporium deacetoxycephalosporin C
XX CC synthetase/hydroxylase polypeptide. The DNA sequence is useful for
XX CC production of cephalosporin in a Penicillium host cell. It is useful for
XX CC the construction of expression vectors for improving the efficiency and
XX CC yield of fermentation involving a wide variety of penicillin and
XX CC cephalosporin antibiotic-producing organism. The expression vector is
XX CC useful in constructing strains for use by the pharmaceutical industry and
XX CC to introduce cephalosporin synthesizing activities into high-level
XX CC penicillin producing Penicillium strain. The DNA sequence is useful for
XX CC screening genomic libraries of organisms that produce cephalosporin C or
```

CC similar compounds for the presence of the gene. It is useful for
CC preparing labelled probes that are used to find expense-encoding DNA
CC sequences in lactam-producing species. (Updated on 11-SEP-2003 to
CC standardise OS field)

XX Sequence 1511 BP; 329 A; 456 C; 429 G; 297 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	9,01e-91	Length:	1511
Score:	933.50	Matches:	178
Percent Similarity:	70.83%	Conservative:	43
Best Local Similarity:	57.05%	Mismatches:	90
Query Match:	57.06%	Indels:	1
DB:	4	Gaps:	1

10719236-1_232-1164 (1-311) x AAF58690 (1-1511)

```

QY 1 MetAspThrThrValProThrPheSerLeuAlaGluLeuGlnGlyLeuHisGlnAsp 20
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 ATGACTTCGAAGTCCCGCTTCTGCTCGACGACTCAAGACGGCAGAGTCTTCCACC 460
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 21 GluPheArgArgCysLeuArgAspIlyGlyLeuPheThrLeuThrAspCysGlyLeuThr 40
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 GAGCTCGCCGAGCCGCTCACACCAAGGGATCTTCTACTTGCAGACGCGCGCTGCTC 520
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 41 AspThrGluLeuSerAlaLysAspIleValIleAspPhePheGluHisGlySerGlu 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 521 GACGACGACACACCTCGCGCGCTGAGACGTGCGACTTTTTCAGAACGAGACGGAG 580
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 AlaGluLysArgAlaValThrSerProValProThrMetArgGlyPheThrGlyLeu 80
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 581 GAGGAGAAAGGGCCCTGACCTCGCCGACCGTAAGCCCGCGCGCTTCTCTGCGCTC 640
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerThrSerAspTyrSerMetCys 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 641 GAGTGGAGAGACACCCCGCTCGTCAACGAGACGGGCAAGTACTCGGACTACTGACGTGC 700
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTyrThr 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 701 TACTCATGGGATCGCGCGCAACCTGTTCCGAAACGGGGGCTTCGAGGACGTCTGGCAG 760
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 761 GACTACTTGCACCGCATGTACGGCGCAGCAAGAGATGTGGCGCGCGCTTCTCAACTCT 820
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 141 ThrGlyThrGluProAspGly---GlyValGluValPheLeuAspCysGluProLeuLeu 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 821 GTGGGCGCCCGCTCGCGGAGAGACATTGATGACTTGTGCGAGTGCATCCCTCTCTC 880
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 160 ArgPheArgTyrPheProGlnValProGluHisArgSerAlaGluGluGlnProLeuArg 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 881 CCGCTACGGTACTTCCCGAAGTCCGAGGAGACCGGTGCGCAAGAGAACCCCTCTCCG 940
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 180 MetAlaProHisTyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsn 199
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 941 ATGGGACCCCACTACCACTATTCGACCATCAGCTGTGTCACACGACACCTCTCCGCAAC 1000
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 200 GlyPheValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrArgPro 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1001 GCGCTTGTGAGCGCTGACGTGCAAGGTGAGGAGGAGATGTGTGACCTCCCGACGCTCC 1060
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 220 AspAlaValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyGlnValLys 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1061 GCGCATGATGCTCTTCTGCGCGCGGTGCGACCTTGCACCGCGCGCGCAAGGTCAAG 1120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 240 AlaProArgHisIleValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThr 259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1121 GGGCCCAACACCGCGGTCAAGTCTCCGCGCGGACGACACCGGTGCGCAGCGCGCAG 1180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 260 SerSerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArg 279
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1181 TCGAGGCTCTTCTTCTGCGCGCGCAAGCCGACTTCAAGCTTCAACGTGACGACGAGTCAAG 1240

```

QY 280 GluCysGlyPheAspValSerLeuAspGlyIleThrAlaThrPheGlnAspTyrIleGly 299

Db 1241 GAGTGGGCTTTCACATCCGATCCCTCGAGACGACGACTTCAAGGAGTGGCTTGGC 1300

QY 300 GlyAsnTyrValAsnIleArgArgThrSerLysAla 311

Db 1301 GCGAATATGTCAACATCGGAGGATTAAGCCGGCG 1336

RESULT 14

ADM41691/C

ID ADM41691 standard; DNA; 4623 BP.

AC ADM41691;

DT 17-JUN-2004 (first entry)

DE Acremonium chrysogenum cephalosporin C biosynthetic cefEF and cefG genes.

OS Acremonium chrysogenum.

FX Key Location/Qualifiers

FT CDS complement (122..1118)

FT /*tag= a

FT /*gene= "cefEF"

FT /*tag= b

FT /*gene= "cefG"

FT /*note= "Contains introns"

FT intron 2618..2696

FT /*tag= C

FT intron 3044..3108

FT /*tag= d

PN WO2004026902-A1.

PD 01-APR-2004.

PF 16-SEP-2003; 2003WO-EP010289.

PR 17-SEP-2002; 2002AT-00001397.

PA (SANO) SANDOZ GMBH.

PI Kuernsteiner H, Friedlin E;

DR WPI; 2004-295383/27.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 4623 BP; 1040 A; 1284 C; 1319 G; 980 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.25e-90	Length:	4623
Score:	933.50 <td>Matches:</td> <td>178 </td>	Matches:	178
Percent Similarity:	70.83%	Conservative:	43
Best Local Similarity:	57.05%	Mismatches:	90

Novel Acremonium chrysogenum protein useful in synthetic or semi-synthetic production of cephalosporin C or its derivatives with antibiotic properties.

Example 6; SEQ ID NO 7; 43pp; English.

The present sequence is that of a genomic DNA region of Acremonium chrysogenum, flanked by XbaI and SmaI BfrI sites, and containing the cephalosporin C biosynthetic cefEF and cefG genes. The invention relates to nucleic acids ADM41686-ADM41688 encoding a novel protein ADM41685 involved in cephalosporin C biosynthesis. It is an object of the present invention to provide a nucleic acid and vectors which code for the new protein and which can be used for transformation of an A. chrysogenum host cell such that the host cell is capable of producing cephalosporin C in good yield. The vector may additionally comprise at least one of the pcDNA, pcBC, cefD1, cefD2, cefEF or cefG genes.


```

Db      187 GAGAAAGCGCGCTGACACGACGTCCGACCATGCGCGCGCTACTCCGCGCTGAG 246
QY      82  SerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCysTyr 101
Db      247 GCCGAGAGACCGCCGACGAGTGCACGAGACCGGTTCTTACACGACATCACTGATGCTTTC 306
QY      102 SerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTPThrgln 121
Db      307 CC-ATGGGATCTCGGGCAACGCTTCCCTCCCGGAGTTCAGCGGGGTGACGAG 365
QY      122 TyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAlaThr 141
Db      366 TACTTCGACAGCTCAACCGCGCGGCGGAGACGCGCGGTGCTGACCGCGAGC 425
QY      142 Gly-----ThrgluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeu 158
Db      426 GCGGCTATAGCGCGGATCGTGGAGCTGAGACGCTGAGACGCTGAGACCGCCGCTG 485
QY      159 LeuArgPheArgTyrPheProGlnValProGluHisArgSerAlaGluGluGlnProLeu 178
Db      486 CTGCGGCTGCGTACTTCCCGAGGTGCGGAGCACCGGTCCCGGACACGACCGCGCC 545
QY      179 ArgMetAlaProHisTyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAla 198
Db      546 CGGATGCGCCGACCTACGACCTGTGATCATCACTTCATCCACGACGCGGTGCGCC 605
QY      199 AsnGlyPheValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrArg 218
Db      606 AACGGCTGTGACCTCCAGCGCGGATCGGCGGAACTGTGAGCTGCGCGCTG 665
QY      219 ProAspAlaValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyGlnVal 238
Db      666 GAGGAGCGCGTGTGATGTCGCGCGGATGCGCCCGCTGCGCACCGAGCGCGCTG 725
QY      239 LysAlaProArgHisIleValAlaAlaProArgArgAspGlnIleAlaGlySerSerArg 258
Db      726 CCCGCGCCCGGACCACTCCGCGTCCCGCGCGGTATGCGCGAGGCGACGACCGC 785
QY      259 ThrSerSerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAla 278
Db      786 AGGTGAGGCTCTTCTTCTGCGCCCGACGACCGACTTCTGTTCTGCGTGGCCAGAGCC 845
QY      279 ArgGluCysGlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrpIle 298
Db      846 CCGAGCTACGCGCTCGACCTCGACATGAGAGACGCGCACTTCGCGCACTGATC 905
QY      299 GlyGlyAsnTyrValAsnIle 305
Db      906 GGCACCACTACGTCAACATG 926

```

Search completed: September 19, 2005, 15:36:01
 Job time : 638 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: September 19, 2005, 15:23:03; Search time 3691 Seconds

(without alignments)
3207.258 Million cell updates/sec

Title: 10719236-1_232-1164

Perfect score: 1636
Sequence: 1 MDTVPTFSLAEIQGLHOD.....ATFDWIGWYVNIKRTSKA 311

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame.p2n.model -DEV=xlh
-O=/cgn2.1/USPTO_spool/h/10719236-1TRANS/runat_19092005_153157_7751/app_query.fasta_1.45
-DB=EST -QFMT=fastap -SUFIX=p2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCLALIGN=200 -THR SCORE=Pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=Pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=10719236-1TRANS_@CGN_1_13437_@runat_19092005_153157_7751 -NCPU=6
-ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_ges1:*
9: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	239.5	14.6	656	CF066872	CF066872 Ac5396 Am
2	187	11.4	710	CF066454	CF066454 Ac4235 Am
3	159	9.7	732	BM407972	BM407972 EST582299
4	156	9.5	1382	AY104706	AY104706 Zea maye
5	152.5	9.3	1101	CI960479	CI960479 Ostricc004
6	146	8.9	632	BM166131	BM166131 BM166131
7	141.5	8.6	1044	CI960315	CI960315 Ostricc004
8	140.5	8.6	1107	CI978343	CI978343 Ostricc031
9	135.5	8.3	1059	CI963525	CI963525 Ostricc009

10	135.5	8.3	1205	7	CK167073	CK167073 FGAS05135
11	135	8.3	768	5	BM141113	BM141113 BM141113
12	133.5	8.2	670	5	BS0519124	BS0519124 EST626539
13	133.5	8.2	928	7	CV289688	CV289688 aof01-2ms
14	133	8.1	773	5	BM072029	BM072029 BM072029
15	133	8.1	778	5	BM117875	BM117875 BM117875
16	133	8.1	779	5	BM023309	BM023309 BM023309
17	133	8.1	832	5	BM139402	BM139402 BM139402
18	132.5	8.1	642	5	BM149619	BM149619 NF107A05F
19	132	8.1	568	9	TA233D09P	TA233D09P
20	131	8.0	721	5	BO858125	BO858125 QG89G12.Y
21	130.5	8.0	562	2	BE776101	BE776101 MY-10-H-0
22	130	7.9	558	4	BS521872	BS521872 15-54 Ste
23	130	7.9	882	7	CV288986	CV288986 aof01-16m
24	129.5	7.9	1023	9	CI982538	CI982538 Ostricc048
25	129	7.9	749	7	CF666703	CF666703 RTCTT1_25
26	128.5	7.9	805	7	CF666848	CF666848 RTCTT1_26
27	128.5	7.9	818	7	CO367625	CO367625 RTT1_35 B
28	128.5	7.9	922	7	CP514629	CP514629 CABU0005
29	128.5	7.9	1065	9	CI978345	CI978345 Ostricc031
30	128	7.8	602	5	BS0519125	BS0519125 EST626540
31	128	7.8	820	6	CB660961	CB660961 OSJNEB03D
32	127	7.8	1053	9	CI963527	CI963527 Ostricc009
33	125.5	7.7	1098	3	CNS09YTZ	CBX32857 Arabidops
34	125	7.6	821	8	CC407079	CC407079 PUNTIP3TB
35	125	7.6	954	9	CI963528	CI963528 Ostricc009
36	124.5	7.6	718	1	AO611155	AO611155 AO611155
37	124	7.6	874	7	CO164665	CO164665 FL01_49-B
38	124	7.6	981	8	CC336687	CC336687 OGAQF35TV
39	124	7.6	1438	3	CNS0A5LR	BX824519 Arabidops
40	123.5	7.5	720	5	BO743319	BO743319 MHE4102 F
41	123.5	7.5	751	7	CF474443	CF474443 RTW2_20
42	123.5	7.5	1223	7	CK167162	CK167162 FGAS05145
43	123	7.5	668	4	BI929386	BI929386 EST549275
44	123	7.5	816	4	BS414419	BS414419 HVSMEK000
45	122.5	7.5	713	2	BS430816	BS430816 SUN006.B0

ALIGNMENTS

RESULT 1
LOCUS CF066872 656 bp mRNA linear EST 08-DEC-2003
DEFINITION CF066872 656 bp mRNA linear EST 08-DEC-2003
ACCESSION CF066872
VERSION CF066872.1 GI:39544568
KEYWORDS EST.
SOURCE Amphidinium carterae
ORGANISM Amphidinium carterae
Eukaryota; Alveolata; Dinophyceae; Gymnodiniales; Gymnodiniaceae;
Amphidinium.
REFERENCE 1 (bases 1 to 656)
Bachvaroff,T.R., Concepcion,G.T., Rogers,C.R., Herman,E.M. and
Delwiche,C.F.
TITLE Dinoflagellate expressed sequence tag data indicate massive
transfer of chloroplast genes to the nuclear genome
JOURNAL Protist 155 (1), 65-78 (2004)
COMMENT Contact: Charles Delwiche
University of Maryland, College Park
H.J. Patterson Hall, College Park, MD 20742, USA
Tel: 301-405-8300
Fax: 301-314-9082
Email: delwiche@umd.edu
Seq primer: CTCGTGCGCAATCCG.
Location/Qualifiers
1..656
/organism="Amphidinium carterae"
/mol_type="mRNA"
/strain="CCMP 1314"
/db_xref="taxon:2961"
/clone="Ac5396"

FEATURES

source
1..656
/organism="Amphidinium carterae"
/mol_type="mRNA"
/strain="CCMP 1314"
/db_xref="taxon:2961"
/clone="Ac5396"

/clone.lib="Amphidinium carterae"
/note="Vector: modified pBluescript SK+; Site_1: EcoRI;
Site_2: NotI"

ORIGIN

Alignment Scores:

Pred. No.:	4,26e-17	Length:	656
Score:	239.50	Matches:	76
Percent Similarity:	47.64%	Conservative:	35
Best Local Similarity:	32.62%	Mismatches:	77
Query Match:	14.64%	Indels:	46
DB:	6	Gaps:	10

10719236-1_232-1164 (1-311) x CF066872 (1-656)

```
Qy 23 ArgArgCysLeuArgAspLysGlyLeuPheThyLeuThrAspCysGlyLeuThrAspThr 42
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 AGAATTGCGCTGAAGAAACAAGCTCTACTGTATGTCACCAATCACTGCGCGTG----- 55
Qy 43 GluLeuLysSerAlaLysAspIleValIleAspPhePheGluHisGlySerGluAlaGlu 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 56 -----GATATC-----GACGGCTTTGCCAAGGAG 79
Qy 63 LysArgAlaValThrSerProVal-----ProThrMetArgArgGlyPheThr 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 80 GTGAGACGCGTGACAGAGACCTGTGATGATCTCTCCGAGAGAGACGACGCTTTTACT 139
Qy 79 Gly-----LeuGluSerGluSerThrAlaGlu 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140 GCAAGAGGTGTTCCCATCAGTGTCTGAGCGCGCTTTGAAACGATCCGTAGCAAGGTA 199
Qy 89 ThrAsnThrGlySerTyrSerAspTyrSerMetCysTyrSerMetGlyThrAlaAspAsn 108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 ATGGGGCAGGGCACTACTGACTGTGTGCAAAAGTGACCTACAAATCACAGCCAGAAC 259
Qy 109 LeuPhePro-----SerGlyAspPheGluArgIleTPrThrGlnTyrPheAspArg 125
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 260 ATACTTCACACAGACGCGAGCGTAGCTTCAGAGAGCGCTGGAGT-----GGT 307
Qy 126 GlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAlaThrGly----- 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 308 GCCACGCGCGTGGATGATGATCGCCATACGCTGCTTCATTGATGATGGCGAAGCGTTG 367
Qy 143 -----ThrGluProAspGlyGlyValGluAlaPheLeuAspCysGlu-----ProLeu 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 368 AACCTCACTGAGAAATCCCGGGTGGCAGGAAATGTTGTGATGGGAAAGCGCTGGGCAACA 427
Qy 159 LeuArgPheArgTyrPheProGlnValProGluHisArgSerAlaGlu-----GlnPro 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 428 ATGGGTTGCTGCGCTACCTCGAGCGTATCTGAGAAATGGCGCATCGATCGGGGGTCCA 487
Qy 178 Leu---ArgMetAlaProHisTyrAspLeuSerMetValThrLeuIleGlnThrPro 196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 GTAGAACCGCATGGCCCCCATACAGACCTTGCGCTGACCATCGTTGTCACCAACCCCG 547
Qy 197 CysAlaAsnGlyPheValSerLeuGluAlaGluValGlyGlyAlaPheThrAspLeuPro 216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 548 TGGCTTAATGGGTTGTGTGCTACCAAGGCTTGATGA-GAAGANTGGTGCATGTGCCA 606
Qy 217 TyrArgProAspAlaValIleValPheCysGlyAlaIle 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 607 NCAGTCGAGACAGCTTTGTCTGTAACCTTGNAAGGTG 645
```

RESULT 2
LOCUS CF066454 710 bp mRNA linear EST 08-DEC-2003
DEFINITION Ac4235 Amphidinium carterae Amphidinium carterae cDNA clone Ac4235
3' mRNA sequence.
ACCESSION CF066454
VERSION CF066454.1 GI:39544150
KEYWORDS EST.
SOURCE Amphidinium carterae
ORGANISM Amphidinium carterae

Eukaryota; Alveolata; Dinophyceae; Gymnodiniales; Gymnodiniaceae;
Amphidinium.
1 (bases 1 to 710)

REFERENCE

1 (bases 1 to 710)

AUTHORS

Bachvaroff,T.R., Concepcion,G.T., Rogers,C.R., Herman,E.M. and

TITLE

Dinoflagellate expressed sequence tag data indicate massive

JOURNAL

transfer of chloroplast genes to the nuclear genome

COMMENT

Protist 155 (1), 65-78 (2004)

CONTACT:

Charles Delwiche

UNIVERSITY:

University of Maryland, College Park

ADDRESS:

H.J. Patterson Hall, College Park, MD 20742, USA

TEL:

301-405-8300

FAX:

301-314-9082

EMAIL:

delwiche@umd.edu

PLATE:

45 row x 6 column; 2

Seq primer:

CTCGTCCCAATTCG.

FEATURES

source

Location/Qualifiers

1..710

/organism="Amphidinium carterae"

/mol_type="mRNA"

/strain="CCMP 1314"

/db_xref="taxon:2961"

/clone="Ac4235"

/clone.lib="Amphidinium carterae"

/note="Vector: modified pBluescript SK+; Site_1: EcoRI;
Site_2: NotI"

ORIGIN

Alignment Scores:

Pred. No.:	6.48e-11	Length:	710
Score:	187.00	Matches:	72
Percent Similarity:	44.81%	Conservative:	23
Best Local Similarity:	33.96%	Mismatches:	67
Query Match:	11.43%	Indels:	51
DB:	6	Gaps:	11

10719236-1_232-1164 (1-311) x CF066454 (1-710)

```
Qy 23 ArgArgCysLeuArgAspLysGlyLeuPheThyLeuThrAspCysGlyLeuThrAspThr 42
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 73 AGAATTGCGCTGAAGAAACAAGCTCTACTGTATGTCACCAATCACTGCGCGTG----- 123
Qy 43 GluLeuLysSerAlaLysAspIleValIleAspPhePheGluHisGlySerGluAlaGlu 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 -----GATATC-----GACGGCTTTGCCAAGGAG 147
Qy 63 LysArgAlaValThrSerProVal-----ProThrMetArgArgGlyPheThr 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148 GTGAGACGCGTGACAGACCTGTGATGATCTCTCCGAGAGAGACGTCAGCTTTTACT 207
Qy 79 Gly-----LeuGluSerGluSerThrAlaGlu 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208 GCAAGAGGTGTTCCCATCAGTGTGTTGTCGCTTTGGAAGCGAATCCGTGGACGAAG 267
Qy 89 ThrAsnThrGlySerTyrSerAspTyrSerMetCys-----TyrSerMetGly 104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 268 ACAGGGGAGGCGCTACTACAGC-----TTGTGTGCAAAAGTGAACACACCGCAGT 321
Qy 105 ThrAlaAspAsnLeuPhePro-----SerGlyAspPheGluArgIleTPrThrGlnTyr 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 322 GACCTGACATCTTCCACACAGCGCGATGATGATCTGACAGAACCTGAGT----- 375
Qy 123 PheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAlaThrGly 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 -----GGCGCCACGACGCTGGGATGATCATCGCTTACGTCCTGCGCATTTGATGGC 429
Qy 143 -----ThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluPro 157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 430 GAAGCCCTGAGCTGACGAGGAATGCCAGGTGGCAAGAAATGTTTGTATGGGAAGCCG 489
Qy 158 Leu-----LeuArgPheArgTyrPheProGlnValProGluHisArg-----SerAla 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 490 TTGGCAACATGCGTTTCTGCGCTACCTGACGATCTGTGANGACCGGCATGTATGTCT 549
```


QY	174	GIUGIUGIunProLeuauXMeAlaAbrohisTyrAspIleuSerMeValThreuleIn	193
DB	550	GGGGGTCCAGTGGCCGGAITGGCCCCGATCAAGCCTTGGCGTATCACTGTGGAC	609
QY	194	GIunThProCYsAlaAsnGlyPheValSerIeuGln	205
DB	610	CAGACCCCGTGGCTAT-GGGTTTGTGTGCTACAA	644
RESULT 3			
LOCUS	BM407972		
DEFINITION	EST582299 potato roots Solanum tuberosum cDNA clone CPRO32H16 5'		
ACCESSION	BM407972		
VERSION	BM407972.1		
KEYWORDS	EST.		
SOURCE	Solanum tuberosum (potato)		
ORGANISM	Solanum tuberosum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Solanum.		
AUTHORS	1 (bases 1 to 732) van der Hoeven,R., Sun,H., Karamycheva,S.A., Tsai,J., Van Aken,S., Uterback,T., Chleming,A., Bougiti,O., Buell,C.R., Romling,C., Tanksley,S. and Baker,B. Generation of ESTs from potato roots unpublished (2001)		
TITLE	Contact: Robin Buell		
JOURNAL	The Institute for Genomic Research		
COMMENT	9712 Medical Center Dr, Rockville, MD 20850, USA Email: potato-array@igr.org This clone can be obtained from the University of Arizona Genomics Institute. Orders can be made through URL: http://genome.arizona.edu/orders/ Seq primer: T3		
FEATURES			
Source	Location/Qualifiers 1..732 /organism="Solanum tuberosum" /mol_type="mRNA" /cultivar="Kennebec" /db_xref="taxon:4113" /clone="CPRO32H16" /tissue_type="roots" /dev_stage="in vitro grown stem cuttings" /lab_host="SOLR" /clone_idb="potato roots" /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; supplier: Cornell University, Tanksley lab; sequencing: The Institute for Genomic Research. Roots were isolated from in vitro grown stem cuttings on CM medium. Roots were isolated two weeks after placing the stem cuttings from in vitro grown plants on medium."		
ORIGIN			
Alignment Scores:			
Pred. No.:	1.26e-07	Length:	732
Score:	159..00	Matches:	63
Percent Similarity:	41.02%	Conservative:	42
Best Local Similarity:	24.61%	Mismatches:	111
Query Match:	9.72%	Indels:	40
DB:	4	Gaps:	11
10719236-1_232-1164 (1-311) x BM407972 (1-732)			
QY	20	AspGluPheArgGlyLeuArgAspIleuPheTyrLeuThrAspCysGly---	38
DB	38	GATACAAATTAGTAAAGCCTGTGTGAGATGCTTTTCCAAATTGATCATCGTGTCTA	97
QY	39	-----LeuThrAspThrGluLeuIleuysSerAlaIysAspIleValIleAspPhe	55
DB	98	CCTTTGTATCTCAGTCAAGCTCTTAAAGGTGCCAAG-----TCCTTCTTT	145

OY		56	GIUHLISGLSERGILUALIGLULYAARGALAVAlHLSerProValProthr-----Met	73
Db		146	GAAATCTCCAGAGAACCAACTCAATAGT-----TCTCCACGTCCATAATGAATGTCGT	199
OY		74	ArgArGSLyPheThrngLYLeuGlSerLUserThraIagInLIeThraSnthrGLyser	93
Db		200	CCTGGAGGCCTTAACAAGAACCCCTAATCCTTTAT-----GAG	238
OY		94	TyrSerApLyTySerMetCyFTyrsErMetGLyThrAlaaPseuLeuPheProSeGLy	113
Db		239	TTTTAGTGAAATTCCTTAATTATGCTTCCACCTGGTCTCGACTTCATATCTTCCACAGCAAT	298
OY		114	-----AsPheGluAaGLIetTrpThgLnItyrPheAspaArgLnItyrThAlaserArg	131
Db		299	CCTCCTCAATTTCCAGAGAGTGAGTAGAAGATTATCTCCCAATTCATGAAGATGGAATG	358
OY		132	AlaValAlaArgGluValleuArgAlathrcGLyThrcGluProAspGLyGLValGLuaLa	151
Db		359	CTGTGTCGAAGAATCCCTTAAGGAGAGTGTCTGGGCTT---CCTCCCTGTGTTCTTGAAGA	415
OY		152	Phe-----LeuAspCySGluProLeuLeuArgPheArgTyrrPheProGlnValpro	168
Db		416	TTCAACAATGACAGAGAGCTGGAGTTCTCATCTCTTAATTATTACTTGCCG-----	466
OY		169	GLUHIsarGSeralIaGLUGlUGlnProLeuArgMetAlaproHisTyrrAspLeuSerMet	188
Db		467	-----GCACAGAGAAAAGAAAGATGATCAATCCACATMAAGATTTCAGTTC	517
OY		189	ValThrLeuIleGlnInThrProCYalalaSncGLyPheValSerLeuGlnAlaGLIval	208
Db		518	TTTCCAAATGTTTAAACA-----AATGAAGTTGCAAGCCTCGAGGTTCAAAG	565
OY		209	GLyGLyAlaPheThrAspLeuProTyrrArgProAspAlaValleuValPheCySGlyAla	228
Db		566	GAGCGCGCAATGCATCCATATGCCCAACACCAATGATGCCCTGTGTCAACATTGGAT	625
OY		229	IleaIatHrleuValIThrGLyGLnValIyaAlaProArghISHlsvalAlaIaPro	248
Db		626	GCTCTTCAGGTCCTTAAGCATGACAAATTCACAGTGCACGCCCATTAAGTATGTGAGACA	685
OY		249	ArgArGraspgInIIeaIaGLySerSerArgThrsErvalPhepe	264
Db		686	AAT-----GGAAGATCAAGGAATTCATTGGCTTC	718
RESULT 4				
	LOCUS	AY104706	1382 bp	mRNA linear HTC 16-OCT-2002
	DEFINITION	Zea mays PC0076122 mRNA sequence.		
	ACCESSION	AY104706		
	VERSION	AY104706.1	GI:21207784	
	KEYWORDS	HTC.		
	SOURCE	Zea mays		
	ORGANISM	Zea mays		
		Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD Clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 1382) Hainey,C.F., Dolan,M., Mao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morante,M. and Tingey,S.V. Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes Unpublished (2002) 2 (bases 1 to 1382) Coe,E.H.		
	JOURNAL	Direct Submission		
	REFERENCE	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA		
	AUTHORS	If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSU, maizegap.org; ZmB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schubler, Iowa State, then clones may be requested from ZmB:		
	JOURNAL			
	TITLE			
	COMMENT			


```

QY 23 ArgArgCysLeuArgAspLysGlyLeuPheThyLeuThrAspCysGlyLeuThrAspThr 42
DB 250 AGGCTCGCCTGGAGACAGTGGGGCTTTCTTCAGGTTGGACCATGAGAGAGAGAG 309
QY 43 GluLeuLysSerAlaLysAspIleValIleAspPhePheGluHisGlySerGluIleGlu 62
DB 310 ACGATGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 369
QY 63 LysAlaGlaValThrSerProValProThrMetAlaArgGlyPheThrGlyLeuGluSer 82
DB 370 GAGAAATGCCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405
QY 83 GluSerThrAlaGlnIleThrAsnThrGlySerThrSerAspThrSerMetCysThrSer 102
DB 406 TACGGCCACAGCCTTGTCTTCTTCACAGCAGCAAGAGCTGATGGTGCACATGCTCGCC 465
QY 103 MetGlyThr-----AlaAspAsnLeuPheProSer-----GlyAsp 114
DB 466 CTCGGCGTCGAGCGCGCCTTCATCGCGCGCCCAACCTCTGCGCGCACAACTCGCGCAAC 525
QY 115 PheGluValGlyIleThrPheGlnThrPheAspArgGlnIleThrAlaSerAlaValAla 134
DB 526 TTTAGTGAAGACCTGAGAGAGATCTCGTGAGATCAAGGAGCTGCGCGCTGCGTGG 585
QY 135 ArgGluValLeuArgAlaThrGlyThrGluProAspGlyGlyValGluAlaPheLeuAsp 154
DB 586 GAGCAGCATCGCGCGCGCTGCGCGCTGCGCGCGCGAGAGCTGAGACGGATGTTGGGAG 645
QY 155 CysGluProLeuLeuArgPheArgThrPheProGlnIleProGlnHisArgSerAlaGlu 174
DB 646 GCAGTGCAGCGCGGTGAGATGAATCTTACCGCGCGCTGCGCGCG-----CGG 693
QY 175 GluGlnProLeuArgMetAlaProHisThrAspLeuSerMetValThrLeuIleGlnGln 194
DB 694 GAGCTGCTGCTGGGCTGAGCCCGCATCCGACGCGAGCGCGGTGACCGCTGCTCCAGAG 753
QY 195 ThrProCysAlaAsnGlyPheValSerLeuGlnIleGluValGlyValAlaPheThrAsp 214
DB 754 GACGCGCGGTTCGCGCGCTGAGGTGCTCGCGGCG-----GCGCGCGCGTGGTGCC 807
QY 215 LeuProThrArgProAspAlaValLeuValPheCysGlyAlaIleAlaThrLeuValThr 234
DB 808 GTTCCATCCCGTCCCGCGCGCTGCTGTCACATGTCGGAGACCTCGAGTGTGCTCAC 867
QY 235 GlyGlyGlnValLysAlaProArgHisHis-----ValAlaAlaProArgArgSerGlnIle 253
DB 868 AATGGCAGGTACAAAGAGCGTGAGACCGCGCGGTGCGCGAGCGAGACGAC----- 921
QY 254 AlaGlySerSerArgThrSerSerValPhePheLeuArgProAsnAlaAspPheThrPhe 273
DB 922 -----CGATGTCGTCGTCACTTCTACGCGCGCGCTGACGACCTCGAGCTC 969
QY 274 SerValProLeuAlaArgGlyCysGlyPheAspValSerLeuAspGlyGluThrAlaThr 293
DB 970 GGC-----CGCTGCC-----GAGCTCTGTCGCGAGCGGAGCGCGCGCAG 1011
QY 294 PheGlnAspThrPheGlyGlyValAsnThrValAlaAsnIleArgArgThrSerIle 310
DB 1012 TACCGAGCTACAAACAGCGAGTACAGCCGCGCACTACGTACACAGCGCG 1062

```

```

REFERENCE 1 (bases 1 to 632)
AUTHORS Satou,Y., Shin-I,T., Kohara,Y. and Satoh,N.
JOURNAL Expressed genes in Ciona intestinalis (2002c)
COMMENT Unpublished (2002)
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satou@ascidian.zoo1.kyoto-u.ac.jp.

FEATURES
source
1..632
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rcincol1b07"
/tissue type="neural complex"
/clone_lib="Nori Satoh unpublished cDNA library, neural complex"

ORIGIN
Alignment Scores:
Pred. No.: 3,36e-06 Length: 632
Score: 146.00 Matches: 38
Percent Similarity: 47.33% Conservative: 24
Best Local Similarity: 29.01% Mismatches: 55
Query Match: 8.92% Indels: 14
DB: 5 Gaps: 3

10719236-1_232-1164 (1-311) x BM166131 (1-632)
QY 159 LeuArgPheArgThrPheProGlnValProGlnHisArgSerAlaGluGluGlnProLeu 178
DB 444 CTCAAATATCTTTATTTATCCACCAATTAAGAGATGTTAAACCGAACAATA----- 391
QY 179 ArgMetAlaProHisThrAspLeuSerMetValThrLeuIleGlnIleThrProCysAla 198
DB 390 GCATGGCGCAACATTCAAGATGGGGTTCATCTTGTCTT----- 349
QY 199 AsnGlyPheVal-----SerLeuGlnIleGluValGlyGlyAlaPheThrAsp 214
DB 348 -----TTTGTTCACAACACTGCGCGGCTTCAGATGAGACGGAAGAACTACAAAGAT 295
QY 215 LeuProThrArgProAspAlaValLeuValPheCysGlyAlaIleAlaThrLeuValThr 234
DB 294 GTCCCTGTAAATGAAGACACGATTTGTATCAACTTGAGATGCGTGGATATTGGACA 235
QY 235 GlyGlyGlnValLysAlaProArgHisHisValAlaAlaProArgArgAspGlnIleAla 254
DB 234 AAAGGAACTAAAGTGCACAAACATCGTAAATATATACCCCAAGATGAATGCGCAAGA 175
QY 255 GlySerSerArgThrSerSerValPhePheLeuArgProAsnAlaAspPheThrPheSer 274
DB 174 AATCATCTCGAAGGTGATGTATCTGCTTATCTGATTAATGATGTTGTATCAT 115
QY 275 ValProLeuAlaArgGluCysGlyPheAspVal 285
DB 114 CAACCTTCGACGTTCAAGGGTGAATGACAGATGTT 82

```

```

RESULT 7
LOCUS CL960315 1044 bp DNA linear GSS 21-SEP-2004
DEFINITION OsIFCC004224 Oryza sativa Expressed library Oryza sativa (indica
ACCESSION CL960315
VERSION CL960315.1 GI:52375327
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group) Embryophyta; Tracheophyta;
Eukaryota; Viridiplantae; Streptophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehharctoidae; Oryzaceae; Oryza.

```

REFERENCE	1 (bases 1 to 1044)			
AUTHORS	Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W. and Wang, J.			
TITLE	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis			
JOURNAL	Unpublished (2004)			
COMMENT	Contact: Chen Chen Department of Bioinformatics Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 101300, China Tel: 86-10-80481559 Fax: 86-10-80488676 Email: chenchen@genomics.org.cn Rice genomic sequence. Class: exon-trapped.			
FEATURES	source	Location/Qualifiers		
	1..1044	/organism="Oryza sativa (indica cultivar-group)" /mol_type="genomic DNA" /db_xref="taxon:39946" /clone_lib="Oryza sativa Express library" /note="Oryza sativa exon trapped genomic sequences "		
ORIGIN	Alignment Scores:			
	Pred. No.: 2,34e-05	Length: 1044		
	Score: 141.50	Matches: 61		
	Percent Similarity: 37.45%	Conservative: 39		
	Best Local Similarity: 22.85%	Mismatches: 108		
	Query Match: 8.65%	Indels: 59		
DB:	9	Gaps: 10		
10719236-1_232-1164 (1-311) x CL960315 (1-1044)				
Oy	30	G yleuPherYrleuthrApyCSg yleuThrAptThgluleuYseSerAllyAsp	49	
Db	232	GGGTTCTTCAGAGTTCGCGGCAAGCGGTCGACCGCGCTTCGCGCGCGCTGCAC	291	
Oy	50	l leVal leAsPhePhegluH sg ySerGlu AglulY A A ValTh SerPro	69	
Db	292	GGCCCAAGCATTCTTCGCCCTCCGCTCGCGAAGCGCGCGC-----CGCGC	345	
Oy	70	ValProThrMetArGaRg lyPheThr-----	78	
Db	346	GTCCCGGCAACCGTTCGCGCTACACAGCGCCACGCGACCGTTCGCTCAAGCTC	409	
Oy	79	-----G yleu uSerGluSerThrAlaGlu leThrAsn	90	
Db	406	CCATGAAGAGACCTCTCTTCGCTTCACAGACGCGCGCGCCCGCTGTC--	462	
Oy	91	Thrg ySerTySerAptYrSerMetCyS YrSerMetG YThrAlaAspAsn leuPhe	110	
Db	463	-----GCCGACTAC-----TTCTCAGACACCTCGGCGCCGACTTCGCG	501	
Oy	111	ProSerG yA pRhegluA g leT pThrGlnT YrPheAspA yGlnT YThrAlaSer	130	
Db	502	CCATGGAGG-----AGGCTGTCACGAAGTACTGCGAAGATGAAGAGACTGCG	553	
Oy	131	ArgAlaValAlaArG uVal leuA A ArgAlaTh G YThrG uProAspG y ValG u	156	
Db	553	CTGACGATCATGCAACCTCGAGAGCTGAGCGCTCGCGTGAAGCGAGGCTACTATAGGAG	612	
Oy	151	AlaPhe leuApyCSg uP oleu leuA yPheA yT YrPheProG uVal ProG uH is	170	
Db	613	TTCTTCGGGCAAGCAAGCTCATGCGGTGCACTACTACCTCCCAATGCCCGAG--	663	
Oy	171	ArgSerAlaG uG uG uP oleuA yMetAlaProH sTyAsp leuSerMetValThr	196	
Db	670	-----CCGAGCGGAGCGCTCGGACGCGGCCGCACTGGGACCCACGCGCTTACC	726	
Oy	191	leu leG uG uThrProCyS AlaAsnG yPheValSer leuG uAlaG uValG y G y	210	

Db	721	ATCCTC-----	-----	CTCCAGACGAGCTGGCGGC	747
Qy	211	-----AlapherthAspleuProTyArq	-----	ProAspAlaValLeu	223
Db	748	CTCAGAGCTCCTGCTGCGACGGGAGATGGCGCCCGCCAGCCCGTCCCGGCGCCATGGTC			807
Qy	224	ValPheCysGlyAlaIleAlaIleAlaIleuValThrGlyGlnValValAlaAlaProArghis			243
Db	808	ATCAACATCGCGCGACACCTTCATGCGCGCTGCGAACGGGAGATATMAGAGCTGCTGCAC			867
Qy	244	HisValAlaIleAlaProArqArgAspGlnIleAlaGlySerSerArgThrSerSerValPhe			263
Db	868	AGCGCGCTGCTGCGACCGCGCGGAGC-----	-----	CGCGCTCGTGGCGCTTC	912
Qy	264	PheLeuArgProAsnAlaAsp	270		
Db	913	TTCTCTGTGCTCCGCGGAGAGAC	933		
RESULT 8					
LOCUS	CL978343	1107 bp	DNA	linear	GSS 21-SEP-2004
DEFINITION	OS:IFCC031672 Oryza sativa Expressed Sequence Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.				
ACCESSION	CL978343				
VERSION	CL978343.1	GI:52411187			
KEYWORDS	GSS.				
SOURCE	Oryza sativa (indica cultivar-group)				
ORGANISM	Oryza sativa (indica cultivar-group)				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.				
AUTHORS	1 (bases 1 to 1107)				
TITLE	Ma,L., Wang,C.J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,G.K.S., Deng X.W. and Wang,J.				
JOURNAL	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis				
COMMENT	Unpublished (2004)				
	Contact: Chen Chen				
	Department of Bioinformatic				
	Beijing Institute of Genomics				
	Chinese Academy of Sciences, Beijing 101300, China				
	Tel: 86-10-80481559				
	Fax: 86-10-80488676				
	Email: chenchen@genomics.org.cn				
	Rice genomic sequence.				
	Class: exon-trapped.				
FEATURES					
source	Location/Qualifiers				
	1..1107				
	/organism="Oryza sativa (indica cultivar-group)"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:39946"				
	/clone_lib="Oryza sativa Expressed Library"				
	/note="Oryza sativa exon trapped genomic sequences "				
ORIGIN					
Alignment Scores:					
Pred. No.:	3,34e-05	length:	1107		
Score:	140.50	Matches:	69		
Percent Similarity:	41.87%	Conservative:	52		
Best Local Similarity:	23.88%	Mismatches:	117		
Query Match:	8.59%	Indels:	51		
DB:	9	Gaps:	14		
	10719236-1_232-1164 (1-311) x CL978343 (1-1107)				
Qy	5	ValProthrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAspGluPheArg	24		
Db	157	ATTCCTACGCTGCACTCGACGAAGTTGCTGCATTCATGTCGTGACAGACGAGACTGA	216		
Qy	25	CysLeu-----	ArgAspGlyGlyLeuPheTyrlLeuThrAspCysGlyLeuThr	40	
Db	217	AGCTGGGGCTCTGCTGTAGCAATGGGAGATTCTTTGAGCTATCAACCATGGCGGTGA	276		

QY 41 AspThrGluLeuLysSerAlaIleValIleAspPheGluHisGlySerGlu 60
 Db 277 GAGGATGTGATCCGCGACGCGAAGACATAGCCGAGTTCTTACGGCTGCCAATGAG 336
 QY 61 AlaGluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeu 80
 Db 337 ACCAAGAAAG-----GCCGACTCGACGCTGCC-----AGTGGCATC 372
 QY 81 GluSerGluSerThrAlaGluIleThrAsnThrGlySerTyrSerAspTyrSer---Met 99
 Db 373 GAGGATATAGCCAGCGCTTCGTGTGTGTCGACAGCAAGACTGTGACTGCGCTGACATG 432
 QY 100 CysTyrSerMet-----GlyThrAlaAspAsnLeuPheProSerGlyAspPheGlu 116
 Db 433 TTCTACCTGCTTCCTCCGCGCCGCGAGTCACAGAACATG----- 471
 QY 117 ArgIleTyrThrGluIleTyr-----PheAspArgGluIleTyrAla 129
 Db 472 GCCCTCTGGCCCTGCATCCTCCATCCTTCAGGAATTCGATGATCGGCTACCTCGGAG 531
 QY 130 SerArgAlaValAlaArgGluValLeu-----ArgAlaThrGlyThrGluPro 145
 Db 532 ACGCGAGGGGTGCGAGATGCTGTGGAGTTCATGCGAAGACATGCGCGTACGAGCGG 591
 QY 146 AspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArgPheArgTyrPhePro 165
 Db 592 GGGTGGCTGCTGTGAGGTTCCAGAGCCAGCGGCGATCATGAAACACATACCCG 651
 QY 166 GluValProGluHisArgSerAlaGluGluGlnProLeuArgMetAlaProHisTyrAsp 185
 Db 652 CCGTCCCGGAG-----GCCGCAAGGTGGTGGCTTCGCCGCGACACCGAC 699
 QY 186 LeuSerMetValThrLeuIleGlnIleThrProCysAlaAsnGlyPheValSerLeuGln 205
 Db 700 GCCGCTCGCTGACCTGCTCTCTGAG-----GTGAACAGAGTCCGCGGCTGAG 750
 QY 206 ---AlaGluValGlyValaPheThrAspLeuProTyrArgProAsp-----AlaVal 222
 Db 751 ATCAGAGACCCCGCGCGAGGTGCTCGCGCGCCCTCCGACGACGCGCGCTTC 810
 QY 223 LeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyValValLeuAlaProArg 242
 Db 811 ATCGTCAACGTGCGGACATCTCGAGATCATGACAGCAAGCAAGTACCGAGCGTGGAG 870
 QY 243 HisHis-----ValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSer 261
 Db 871 CACAGCGCGGTGTCGCGCGCGAGGAG-----CGCGTGTGCGCG 912
 QY 262 ValPhePheLeuArgProAsnAlaAsp 270
 Db 913 GCGGTGTTCACACGACCGCTGCCAGAC 939
 RESULT 9
 LOCUS CL963525 1059 bp DNA linear GSS 21-SEP-2004
 DEFINITION OsRFLCC09656 Oryza sativa Expressed Library Oryza sativa (indica
 cultivar-group) genomic, genomic survey sequence.
 ACCESSION CL963525
 VERSION CL963525.1 GI:52381774
 KEYWORDS GSS.
 SOURCE Oryza sativa (indica cultivar-group)
 ORGANISM Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 REFERENCE 1 (bases 1 to 1059)
 Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
 Zhao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
 Wong, G. K. S., Deng, X. W. and Wang, J.
 Analysis of transcriptional regulation of the rice genome and
 its comparison to Arabidopsis
 JOURNAL Unpublished (2004)

COMMENT Contact: Chen Chen
 Department of Bioinformatic
 Beijing Institute of Genomics
 Chinese Academy of Sciences, Beijing 101300, China
 Tel: 86-10-80481559
 Fax: 86-10-80488676
 Email: chenchen@genomics.org.cn
 Rice genomic sequence.
 Class: exon-trapped.
 FEATURES
 source
 1. 1059
 Location/Qualifiers
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="genomic DNA"
 /db_xref="taxon:39946"
 /clone_lib="Oryza sativa Expressed Library"
 /note="Oryza sativa exon trapped genomic sequences"
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.00012 Length: 1059
 Score: 135.50 Matches: 81
 Percent Similarity: 37.95% Conservative: 34
 Best Local Similarity: 26.73% Mismatches: 127
 Query Match: 8.28% Indels: 62
 DB: 9 Gaps: 13
 10719236-1_232-1164 (1-311) x CL963525 (1-1059)
 QY 5 ValProThrPheSerLeuAlaGluLeu-----GlnGln 15
 Db 151 ATCCCAACATAGACGAGGCGACTGAGCCGCGAGCGGTGACGAGTGTGACGAC 210
 QY 16 GlyLeuHisGlnAspGluPheArgArgCysLeuArgAspGlyLeuPheTyrLeuThr 35
 Db 211 GCGCGGAGCGCGGCGGCGCTCGGCGCGCTTCAGTGTGGGCGCTTCTGCGGAGCC 270
 QY 36 AspCysGlyLeuThrAspThrGluLeuLysSerAlaIleValIleAspPhe 55
 Db 271 GCGCAGGCGATGCGGAGCGCTTCTCGACGAGTCTCGCGCGGACCGGGAATTCTTC 330
 QY 56 GluHisGlySerGluAlaGlyLysArgAlaValThr---SerProValProThr----- 72
 Db 331 CACCTCCGCGCGGAGGAGAGAGAGATGACAAACGTGTGCGCGCGCGAGCGCGAC 390
 QY 73 -----MetArgGlyPheThrGlyLeuGluSerGluSerThrAla----- 86
 Db 391 GCGCTGCGCGCGCGCGCGAGGTTCCAGCGCTACGAGCTACCGGATCGACCGGTGAC 450
 QY 87 -----GlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys----- 100
 Db 451 ACCAGACGACGATCTCGACTGTGAGCCGAGCTTACCTCCAGGTGACCGCGAGAG 510
 QY 101 -----TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArg 117
 Db 511 GAGCGCGGCTGAGATTCTGCGCGAGACCGCGCG---CGTTCGCGGAGCTCTCGAG 567
 QY 117 GlnTyrThrGlnIleThrAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluVal 137
 Db 568 GAG-TACACCGCG-----CGAGCGAGCAGAGGTGTTCCGCGCGGT 605
 QY 137 IleuArgAlaThrGlyThrGluProAspGlyValGluAlaPheLeuAsp----- 154
 Db 606 GCTGCGCGGCGCGCGAGGTGCTCGGGGTTCGGGAGAGAGTTCCTTCGCGCAAGGTGCG 665
 QY 155 -----CysGluProLeuLeuArgPheArgTyrPheProGluValProGluHisArgSe 172
 Db 666 CGAAGAGTGACGACGTACGCGCGTTCACCTACTACCGCGCATGCGCGCG----- 717
 QY 172 AlaGluGluGlnProLeuArgMetAlaProHisTyrAspLeuSerMetValThrLeuIle 192
 Db 718 ---CCGAGGTGTGTGAGCGCTCAAGCCCAACGACGACAACTCGCTTCACCGCTTC 773
 QY 192 eGlnIleThrProCysAlaAsnGlyPheValSerLeuGlnAlaGluValGlyAlaPhe 212

Db 774 CCTCTGCAGACAGACGTCGGCGGC-----CTCAGCTGCTCAGACGCGCCCTG 824
Qy 212 eThrAspLeuProTyraArgProAspAlaValIleuValPheCysGly----- 227
Db 825 GCTGCAATTCGGGTGCTACCAACGAGAGCTCTGCTGTCGCCGCGCAGCATTTAGATT 884
Qy 228 ----AlaIleAlaThrIleuValThrGlyGlyGlnVal---LysAlaProArgHisIleVal 245
Db 885 GTTTCGATTTGCTTGCTGTGCATCATAGACAAAGTTCATGCGCGCGTGCACAGGCT 944
Qy 245 IAlaAlaProArgHisArgAspGlnIleAlaGlySerSerArgThrSerSerValPhePheLe 265
Db 945 GGTACACGACGACGACGAG-----AGCATTCGGTGTGGATTTCTA 986
Qy 265 uArgPro 267
Db 987 CCAGCCG 993
RESULT 10
CK167073 1205 bp mRNA linear EST 05-DEC-2003
LOCUS FGAS051352 Triticum aestivum FGAS: Talc7 Triticum aestivum cDNA,
DEFINITION mRNA sequence.
ACCESSION CK167073 GI:39000780
VERSION EST.
KEYWORDS Triticum aestivum (bread wheat)
SOURCE Triticum aestivum
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 1205)
Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,
Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,
Peniket, C., Roach, J.L. and Sarhan, F.
Functional Genomics of Abiotic Stress in Wheat and Canola Crops
Unpublished (2003)
Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas.estecs.usask.ca
This sequence is the direct result of the Base calling software
Phred (default parameters). It is the raw base calls. To aid in the
identification of the high quality insert the software Lucy
(default parameters) has been run on this sequence. Lucy identified
the region [60,850].
Plate: Talc711 row: G column: 16.
location/Qualifiers
1..1205
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="wheat line CI 14106"
/db_xref="taxon:4565"
/lab_host="DH5 alpha"
/clone_lib="Triticum aestivum FGAS: Talc7"
/note="Organ: Crown; Vector: pGEM-T; SSH (suppression
subtractive hybridization) cDNA library from genotype
CI14106 cold hardened at 2 C for 21 days and 49 days
(equal amount of cDNA pooled together before subtraction,
tester) and subtracted against genotype CI14106
non-hardened (20 C) (driver). Nitro-pyrole anchored
oligo-dT priming and non-directional cloning."

Alignment Scores: 0.000145 Length: 1205
Pred. No.: 135.50 Matches: 71
Score:

Percent Similarity: 38.44% Conservative: 52
Best Local Similarity: 22.19% Mismatches: 122
Query Match: 8.28% Indels: 75
DB: 7 Gaps: 14
10719236-1_232-1164 (1-311) x CK167073 (1-1205)
Qy 4 ThrValProThrPheSerLeuAlaGluLeu-----GlnGlnIleuHisGln 19
Db 259 AGCATTCCTGTGATGACCTTGCTGAGCTCTTGATCCCGAGCTATACGCTTGAAGCG 318
Qy 20 AspGluPheArgArgCysLeuArgAspGlyLeuPheTyrlleuThrAspCysGlyLeu 39
Db 319 GCCAGGCTCAGGTCGGCATGTGAGAGTGGGGCTTCTTCAGGTGAAAAACATGAATT 378
Qy 40 ThrAspThrGluLeuLysSerAlaLysAspIleValIleAspPheGluHisGlySer 59
Db 379 CCTGACACAGTTCTTGAATAACATGAGAAAGCACTTCGACACTTCTTCAGGCTTCCCTC 438
Qy 60 GluAlaGluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGly 79
Db 439 GATGAGAGAGACGATTTGGCCAGTTGCTCGAGACCTCCAGGCTATGAGCCAGCATTT 498
Qy 80 LeuGluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrlSerAspTyrlSerMet 99
Db 499 GTTGAGTCCGAACAT----- 513
Qy 100 CysTyrlSerMetGlyThrAlaAspAsnLeuPhe-----ProSerGlyAspPhe 115
Db 514 ---CAAAGCTCAGCATGCTGTGACAGGCTTTCCTAGTACCCAGCCAGCCTCAGATCGT 570
Qy 116 Glu---ArgIleThr-----ThrGlnTyrlPheAspArgGlnTyrlThrAlaSerArgAla 132
Db 571 GAAATAGACCATGCGCTTCAACTCCTGTACAGTTAGAGAA-----TCCATGAGTGG 624
Qy 133 ValAlaArgGluValLeuArgAlaThrGlyThr-----GluProAspGly 147
Db 625 TACTCCTCGAGCTGATGAGGAGGAGTGGCTGGCTCATGGAATCATCGCTGAAGACCTA 684
Qy 148 GlyValGluAlaPheLeuAsp-----CysGluProLeuLeuArgPheArgTyrlPhePro 165
Db 685 GGGGTGATGATCCCGCATGATATCATGATGATGAGGATGAGTGAAGTCACTACCT 744
Qy 166 GluValProGluHisArgSerAlaGluGlnGlnProLeuArgMetAlaProHisTyrlAsp 185
Db 745 GCATGCCCC-----GTGGCTCATGATTAAGCTTTGGGATTTCAACACATTCGCAC 795
Qy 186 LeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGlyPheValSerLeuGln 205
Db 796 ATTTCATGCTAACGCTGATGAGAGTGAATATGTTGGAGGCTCCANGATCAGAGA 855
Qy 206 AlaGluValGlyLysAlaPheThrAspLeuProTyrlArgProAspAlaValIleuValPhe 225
Db 856 NCAGAT-----GCATGGGTGCGCGGTAAACCCACCCCAAGGCATTTGGTGAAT 906
Qy 226 CysGlyValAlaIleAlaThrIleuValThrGlyGlyGlnValIleuValAspProArgHisIleVal 245
Db 907 GTTGGTGACTTTCTTGAGATTAATGACAAACGAAATTAACAAAGCTTGAAGCACAGAGTC 966
Qy 246 ---AlaAlaProArgHisArgGlnIleAlaGlySerSerArgThrSerSerValPhePhe 264
Db 967 ACCGNTATCTCAGACGAGGAGATGATCAATTTCTGCT----- 1005
Qy 265 LeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCysGlyPheAsp 284
Db 1006 -----TTTCACCTTCCA-----AACTTGAAT 1026
Qy 285 ValSerLeuAspGlyGluThrAlaThrPheGlnAspTyrlIleGlyLysAspTyrlValAsn 304
Db 1027 ATGAGCTTG-----GGCAATATCTGAAT 1050
RESULT 11
BM141113/c

LOCUS BW141113 768 bp mRNA linear EST 03-NOV-2002
 DEFINITION BW141113 Nori Satoh unpublished cDNA library, gastrula and neurula
 Ciona intestinalis cDNA clone rcign058c02 3', mRNA sequence.
 ACCESSION BW141113
 VERSION BW141113.1 GI:24498338
 KEYWORDS EST.
 SOURCE Ciona intestinalis
 ORGANISM Ciona intestinalis
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 Phlebobranchia; Clonidae; Ciona.
 REFERENCE 1 (bases 1 to 768)
 Satou, Y., Shin-I, T., Kohara, Y. and Satoh, N.
 TITLE Expressed genes in Ciona intestinalis (2002c)
 JOURNAL Unpublished (2002)
 COMMENT Contact: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@sci.kyoto-u.ac.jp.
 FEATURES
 source
 1..768
 Location/Qualifiers
 /organism="Ciona intestinalis"
 /mol_type="mRNA"
 /db_xref="taxon:7719"
 /clone="rcign058c02"
 /cissue_type="whole body"
 /dev_stage="gastrula and neurula"
 /clone_lib="Nori Satoh unpublished cDNA library, gastrula and neurula"
 ORIGIN
 Alignment Scores:
 Pred. No.: 8,61e-05 Length: 768
 Score: 135.00 Matches: 62
 Percent Similarity: 36.24% Conservative: 42
 Best Local Similarity: 21.60% Mismatches: 92
 Query Match: 8.25% Indels: 92
 DB: 5 Gaps: 11
 10719236-1_232-1164 (1-311) x BW141113 (1-768)
 QY 31 LeuPheTYLeuThraSPcYgLYleuThraSPThrGluLeuLYSerAlaLYsPle 50
 Db 744 TTTGTTTACCTCTGCTCAATGCTGAATACACACACTTCGTTGACGAGTAATAAGTA 685
 QY 51 ValIleASPhePheGluHISglYSerGluAlaGluLYsArgAlaValThrSerProVal 70
 Db 684 GCTTTGATTTCTTCACAAA---CCTGAAGATGAGAAATTAAAGTTTGCTGAATATGC 628
 QY 71 ProThrMetArgARGLYPheThrGluYeuGluSerGluSerThraLagInleThraSn 90
 Db 627 AACCAATTC---GGAATACGACTCTATTTGGAAGGAAAG---TTAGAT 586
 QY 91 ThrGlySerTYrSerAPYrSerMetCysTYrSerMetGlyThr----- 105
 Db 585 GATGAGAGCCAGGTTGATTAAAGCTTTCATTAAACATGAGCTTGCGACGATTAACGAC 526
 QY 106 -----AlaSPasnLeuPheProSerGlyASPheGluArgIleThrPthr 120
 Db 525 CCATCATTAATGATGCGCGATTAATTAATTCAC----- 493
 QY 121 GlnTYrPheASPARGInTYrThraLaserArgAlaValAlaArgGluValleuArgAla 140
 Db 492 -----CAATTTCTTCACAC----- 478
 QY 141 ThrGlyThnGluProASPGLYglYValGluAlaPheLeu---ASPcYgLUProLeu--- 158
 Db 477 -----GTAATCATTAATGAAGAGATTGCAAAATGCTTCT 442
 QY 158 ----- 158

Db 441 CTCAGATTGTTGATTTCTTATTCATTGGCGCTAAATTAACAGACCAACCAACTATCA 382
 QY 159 -----LeuArgPheArgTYrPhe 164
 Db 381 GCACTTCATGATGATTCATATGTCAAATATCATCTTGTTACGCTGCTTATATAT 322
 QY 165 ProGluValProGluHISArgSerAlaGluGluInProLeuArgMetAlaProHisTYr 184
 Db 321 CCA-----CCTATTATACAGAGAGTTAGAGAAAGGTCAAAATGATGCTTGAACACACC 268
 QY 185 ASPLeuSerMetValThrLeuIleGlnInThrProCysAlaAsnGlyPheValSerLeu 204
 Db 267 GATTATGGCTTAATGACACTTCTCTTCAACAT-----TCTGTTGGTGATTCGAGATC 214
 QY 205 GlnAlaGluValIGLYglYAlaPheThraSPLeuProTYrArgProASPAlaValleuVal 224
 Db 213 AAATCGCA---GATGGACTATATGATCATCTTGTCMAATACAGACACTGTTTGATA 157
 QY 225 PheCYgglYAlaIleAlaThrLeuVal-ThrGlyGLYglYValAlaProArgHisHI 244
 Db 156 AA-TGTGGCAGATTGTTTACAACTGACACATGAGACCTAAATCTTACGCCACATCG 98
 QY 244 sValAlaIaProArgArgArgPGLInleAlaGlySerSerArgThrSerSerValPhePh 264
 Db 97 CATCTGGTTCAGATGATATCAACTAGAGGCGCTACCCAGACATGATTCCTTCTT 38
 QY 264 eLeuArgProASPnAlaASP 270
 Db 37 TGTTCATCTCTTGTGAGAC 19
 RESULT 12
 BOS19124
 LOCUS BOS19124
 DEFINITION BOS19124 670 bp mRNA linear EST 07-MAR-2003
 EST626539 Generation of a set of potato cDNA clones for microarray
 analyses mixed potato tissues Solanum tuberosum cDNA clone STMJ007
 5' end, mRNA sequence.
 ACCESSION BOS19124
 VERSION BOS19124.1 GI:21377993
 KEYWORDS EST.
 SOURCE Solanum tuberosum (potato)
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamiales; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 670)
 Buehl, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C.,
 Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and
 Karamchewa, S.A.
 TITLE Generation of a set of potato cDNA clones for microarray analyses
 JOURNAL Unpublished (2002)
 COMMENT Other ESTs: EST626540
 Contact: Robin Buehl
 The Institute for Genomic Research
 9712 Medical Center Dr. Rockville, MD 20850, USA
 Email: potato-array@tigr.org
 This clone can be obtained from the University of Arizona Genomics
 Institute. Orders can be made through URL:
 http://genome.arizona.edu/orders/
 Seq primer: T3.
 FEATURES
 source
 1..670
 Location/Qualifiers
 /organism="Solanum tuberosum"
 /mol_type="mRNA"
 /cultiVar="Kennebec or Binjle"
 /db_xref="taxon:4113"
 /clone="STMJ007"
 /tissue_type="mixed tissues"
 /lab_host="SOLR"
 /clone_lib="Generation of a set of potato cDNA clones for
 microarray analyses mixed potato tissues"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; supplier: Combination of untreated and Phytophthora
 infestans-treated libraries of scions, leaves, leaflets,

axillary buds of stem explants, petioles, germinating
eyes, tubers, or roots."

ORIGIN

Alignment Scores:
Pred. No.: 0.000106 Length: 670
Score: 133.50 Matches: 55
Percent Similarity: 41.28% Conservative: 42
Best Local Similarity: 23.40% Mismatches: 103
Query Match: 8.16% Indels: 15
Gaps: 30

10719236-1_232-1164 (1-311) x BQ519124 (1-670)

QY 20 AspgluPheargAgyCsleuArgAsplysGlyLeupheTyrlLeuthraSpCySgLy--- 38
Db 38 GATACAAATTAGTAAAGCCGTGTGGAGTATGTTTCCAAATTGGAACATGAGGTGA 97
QY 39 -----LeuthraSpThrgLuleuysSeraAlaYsaApilValIleaspPhaPhe 55
Db 98 CCTTTGATCTCATAGCAAGCTCTAAAGGTGCCAG-----TCTTTCTTT 145
QY 56 GluHisGlySerGluAlaGluLysArgAlaValThrSerProValProThr-----Met 73
Db 146 GAATCTCCACAGAACCAACTCAATGT-----TCTCCAGCTGCTTAATGATCTGTT 199
QY 74 ArgArgGlyPheThrGlyLeuGlySerGlySerThrAlaGlnIleThrAsnThylSer 93
Db 200 CCTGCAGGCTTAAACAGAACCTTAATCTTTCTTAT-----GAG 238
QY 94 TySerAspTySerMetCysTySerMetGlyThrAlaAspAsnLeuPheProSerGly 113
Db 239 TTTAGTGAATTCCTTATATCTTCCACCTGTTCTGACTTCATATCTTCCAGACAAT 298
QY 114 -----AspPheGluArgIleTrpThrGlnIleThrPheAspArgGlnIleThrAlaSerArg 131
Db 299 CCTCTCAATTTCCAGAGGTGATGAGAGTATTCTCCCAATTCATGAGATAGATG 358
QY 132 AlaValAlaArgGluValLeuArgAlaThrGlyThrGluProAspGlyGlyValGluAla 151
Db 359 GTTGTCCAAAGAAATCTTAGCGAGTGTCTGGGCTT---CTCCCTGTGTTCTTGAAGAA 415
QY 152 Phe-----LeuAspCysGluProLeuLeuArgPheArgTyrlPheProGlnValPro 168
Db 416 TTCAACATGACAGAGCTGGGATTTCTCTCATTTTATTTTATCTTGCCG----- 466
QY 169 GluHisArgSerAlaGluGluGlnProLeuArgMetAlaProHisTyrlAspLeuSerMet 188
Db 467 -----GCACACAGAGAAAGAAAGATAGATCAATCCACATTAAGATTTCAGTTGC 517
QY 189 ValThrLeuIleGlnIleThrProCysAlaAsnGlyPheValSerLeuGlnAlaGluVal 208
Db 518 TTCCAAATGTTTACAA-----AATGAAGTTGCGAGCCCTGGAGGTTCAAAA 565
QY 209 GlyIleAlaPheThrAspLeuProTyrlAspProAspAlaValLeuValPheCysGlyAla 228
Db 566 GAGCGCGAATGATCTTAATAGCCCCACACCAATGAGCCCTGTTGTCAACATTTGGAT 625
QY 229 IleAlaThrLeuValThrGlyGlnValIleAlaProArgHis 243
Db 626 GCTCTCAAGTCTTAGCAATGACAAATTCAAGATCCACGCCAT 670

RESULT 13
LOCUS CV289688 928 bp mRNA linear EST 23-SEP-2004
DEFINITION aof01-2ms2-e08 aof01 Asparagus officinalis cDNA clone
ACCESSION CV289688
VERSION CV289688.1 GI:52574247
KEYWORDS EST.
SOURCE Asparagus officinalis (garden asparagus)
ORGANISM Asparagus officinalis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE

1 (bases 1 to 928)
dePamphilis, C., Carlson, J., Ma, H., Soltis, D., Soltis, P.,
Oppenheimer, D., Frohlich, M., Doyle, J., Tankeley, S., Webb, M.,
Leebens-Mack, J., Landherr, L., Iluc, D. and Wall, K.
Generation of ESTs from early male inflorescences of Asparagus
officinalis

JOURNAL

Unpublished (2004)
Contact: Claude dePamphilis or James Leebens-Mack
Mueller Laboratory
Penn State University
208 Mueller Laboratory, Department of Biology, ATTN Rm212, Penn
State University, University Park, PA 16802, USA
Tel: 814 865 6413
Fax: 814 865 9131

COMMENT

Email: cswd3psu.edu or jh110@psu.edu
The sequence provided is trimmed of vector and low quality regions.
Full sequence and original trace file are available from the Plant
Genome Network website (http://pgn.cornell.edu)
Plate: aof01-2ms2 row: e column: 08
Seq primer: M13F.

FEATURES

Location/Qualifiers
1..928
/organism="Asparagus officinalis"
/mol_type="mRNA"
/db_xref="taxon:4686"
/clone="aof01-2ms2-e08"
/tissue_type="male inflorescences"
/lab_host="SOLR"
/clone_lib="aof01"
/note="Vector: pBluescript SK (+/-); Site 1: EcoRI;
Site 2: XhoI; this is a directionally cloned,
non-normalized library. This library has been generated by
the Floral Genome Project (FGP). The Floral Genome Project
is funded by NSF's Plant Genome Research Program
(DBI-0115684). More information about the project can be
obtained at http://fgp.bio.psu.edu"

source

ORIGIN

Alignment Scores:

Pred. No.: 0.00017 Length: 928
Score: 133.50 Matches: 61
Percent Similarity: 38.73% Conservative: 49
Best Local Similarity: 21.48% Mismatches: 127
Query Match: 8.16% Indels: 47
Gaps: 11

10719236-1_232-1164 (1-311) x CV289688 (1-928)

QY 2 AspThrThyValProThrPheSerLeuAlaGluLeuGlnGlnIleuHisGlnAspGlu 21
Db 157 GACATCCCGTCTTGAACCTCTCTGTCGCAC-----CGCGACAGT 198
QY 22 PheArgArgCysLeuArgAspLys-----GlyLeuPheTyrlLeuthraSpCys 37
Db 199 GTGTCAAGCGCATACCCAGCCGCTGATGAGATGGGAGATTTCAGTTGGTGAACAC 258
QY 38 GlyLeuThrAspThrgLuleuysSeraAlaYsaApilValIleaspPhePheGluHis 57
Db 259 GGGATCCCGACGAGGTGATAGAGAGCTGACAGAGGTGGGACGAGATTCTTC----- 312
QY 58 GlySerGluAlaGluLysArgAlaValThrSerProValProThrMetArgArgGlyPhe 77
Db 313 GGGCTGCCGACAGAGAGAGAGGCTCTACGCCACGATCCGGG---TCGGGAGACTTC 369
QY 78 ThGlyLeuGlySerGlySerThrAlaGlnIleThrAsnThylSerTyrlSerAspTyrl 97
Db 370 CAGGGGTAGCGGACTTAAGCTGACAGAGATTTGAGAGGAGAGAGCTTGGGTATAT 429
QY 98 SerMetCysTyrlSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGlu--- 116
Db 430 TTGTTT-----CACACGTTGTGCGGAGAGAGAGGATGATTAAC 468


```

QY 117 ArgIleThrThrGlnIleThrPheAspArgGlnIleThrAlaSerArgAlaValAlaArgIle 136
Db 469 AAGTATTGGCCCAAGAACCTCCCGAGTATAGAAAGGCTAACGAAGATATCCCAAGAC 528
QY 137 ValLeuArgAlaThr-----GlyThrGluPro 145
Db 529 CTCCTTCAGATAGTGCATGAATGCTGAGCAGCTTATCGTCGCTCGATTAAGACCT 588
QY 146 AspGlyGlyValGlnAlaPhe-----LeuAspCysGluProLeuArgPheArgTyr 163
Db 589 AACGCACTAAAGATGACAGTCCGAGCGCATCTCGAGTACTCGTGAAGATCAACTAC 648
QY 164 PheProGlnValProGlnHisArgSerAlaGluGlnProLeuArgMetAlaProHis 183
Db 649 TACCGCCGCTGGCCAGC-----CCCGATCTCGCCCTGCGCTGCGCCGAC 696
QY 184 TyrAspLeuSerMetValThrLeuIleGlnIleThrProCysAlaAsnGlyPheValSer 203
Db 697 ACCGACATGTCCGCCCATCAATCCTC-----GTCCGATGAGTACCGGGG 744
QY 204 LeuGlnAlaGluValGlyValAlaPheThrAspLeuProTyrArgProAspAlaValLeu 223
Db 745 CTGCAAGCTGTCAAGATATCATCTGTTGATGCCAATATATATCCCTAATGCGCTCAAT 804
QY 224 ValPheCysGlyAlaIleAlaThrLeuValIleArgGlyGlnValAlaAlaProArgHis 243
Db 805 GTTCATATCTGGATCAAAATTGATGATCTTAAGCAATGGAAGTACAGAGTGTCTGTCAC 864
QY 244 HisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSerValPhe 263
Db 865 AGAACAGCGCTCAACAGAG-----AAAGCAGATGTCATGCGCAAGG 909
QY 264 PheLeuArgPro 267
Db 910 TTCTGCTCACCA 921

RESULT 14
LOCUS BM072029 773 bp mRNA linear EST 20-OCT-2002
DEFINITION BM072029 Nori Satoh unpublished cDNA library, cleaving embryo Ciona
ACCESSION BM072029
VERSION BM072029.1 GI:24173441
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
Bukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Ciona.
REFERENCE 1 (bases 1 to 773)
AUTHORS Satou, Y., Shin, I. T., Kohara, Y. and Satoh, N.
TITLES Expressed genes in Ciona intestinalis (2002c)
JOURNAL Unpublished (2002)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source
1..773
location/Qualifiers
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rcic1097p04"
/tissue_type="whole body"
/dev_stage="cleaving embryo"
/clone_lib="Nori Satoh unpublished cDNA library, cleaving
embryo"

ORIGIN
Alignment Scores:

```

```

Pred. No.: 0.000149 Length: 773
Score: 133.00 Matches: 62
Percent Similarity: 35.57% Conservative: 44
Best Local Similarity: 20.81% Mismatch: 101
Query Match: 8.13% Indels: 92
DB: 5 Gaps: 11

10719236-1_232-1164 (1-311) x BM072029 (1-773)

QY 20 AspGluPheArgArgCysLeuArgAspIleGlyLeuPheTyrLeuThrAspCysGlyLeu 39
Db 773 GAGGAATAATGCAATGCTTCAAAATGTTGTTTGTCTACCTCGTCAATATCGAATA 714
QY 40 ThrAspThrGluLeuValSerAlaAlaAspIleValIleAspPhePheGluHisGlySer 59
Db 713 CACAGCCTTGGGTGAGAGTAATTAAGTACCTTGGATTTCTTACAAA-----CCT 657
QY 60 GluAlaGluValAlaValAlaThrSerProValProThrMetArgArgGlyPheThrGly 79
Db 656 GAAATGGAATTAATTAATTTGTAAGATTCACAAATTC-----GATATGACACTCT 603
QY 80 LeuGluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMet 99
Db 602 ATTGGAAGGAAG-----TTAGATGATGGAAGCCAGGTGATTAATTAAGCTT 555
QY 100 CysTyrSerMetGlyThr-----AlaAspAsnLeu 109
Db 554 TCATTAAACAGGCTTGTGCGACGTAATACAGCCAGCATTAATGACCGAGATTA 495
QY 110 PheProSerGlyAspPheGluArgIleThrThrGlnIleThrPheAspArgGlnIleThrAla 129
Db 494 GTTCCA-----CAATTTCTTCC 477
QY 130 SerArgAlaValAlaArgGluValLeuArgAlaThrGlyThrGluProAspGlyVal 149
Db 476 ACC-----GTG 471
QY 150 GluAlaPheLeu-----AspCysGluProLeu----- 158
Db 470 AAATCATTTATGAAGATTGCAAAATGCTTCTCTCAGAAATTTGGATTCTTATCATTT 411
QY 158 ----- 158
Db 410 GGGCTAAATTAACAGCAACCAACAATATCAGACCTTCAATGATCTTATGTCGA 351
QY 159 -----LeuArgPheArgTyrPheProGlnValProGlnHisArgSerAla 173
Db 350 AATATCATTTCTGCTTACGCTGCTTATATATCA-----CTATTAACAAGATTA 297
QY 174 GluGluGlnProLeuArgMetAlaProHisTyrArgLeuSerMetValIleThrLeuIleGln 193
Db 296 GAGAAAGCTCAAAATGCTTGAACACACCGATTAATGCGCTTAATGACACTTCTTT 237
QY 194 GlnThrProCysAlaAsnGlyPheValSerLeuGlnAlaGluValGlyAlaPheThr 213
Db 236 CAAGAT-----TCTGTTGGTGAATGAGATCAATGCGA-----GATGGACTTATATC 186
QY 214 AspLeuProTyrArgProAspAlaValLeuValAlaPheCysGlyAlaIleAlaThrLeuVal 233
Db 185 GATCCTTGTCCAAATACAGACACTGTTTGAATAA-TGTGGCAATTTGTTAATAATG 127
QY 234 -ThrGlyGlyGlnValAlaPheProArgHisValAlaAlaProArgAspArgGlnIle 253
Db 126 GACACATGAGACCTTAATTTAGCCACATCGCATCTGTTGTCAGATGATACACTTA 67
QY 253 AlaGlySerSerArgThrSerSerValPhePheLeuArgProAsnAlaAsp 270
Db 66 GAGGGCTAACCCACGACATGATGCTTCTTTGTTGATCTCTGTGAGAC 15

RESULT 15
LOCUS BM117875 778 bp mRNA linear EST 24-OCT-2002
DEFINITION BM117875 Nori Satoh unpublished cDNA library, tailbud embryo Ciona

```

intestinalis cDNA clone rcltb078c2 3', mRNA sequence.
BM117875
GI:24364540
EST.
Ciona intestinalis
Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Ciona.
1 (bases 1 to 778)
Sato, Y., Shin-I, T., Kohara, Y. and Sato, N.
Expressed genes in Ciona intestinalis (2002c)
Unpublished (2002)
Contact: Nori Sato
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: sato@ascidian.zool.kyoto-u.ac.jp.
location/Qualifiers
1..778
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rcltb078c2"
/tissue_type="whole animal"
/dev_stage="tailbud embryo"
/clone_lib="Nori Sato's unpublished cDNA library, tailbud
embryo"
ORIGIN
Alignment Scores:
Pred. No.: 0.00015 Length: 778
Score: 133.00 Matches: 62
Percent Similarity: 35.57% Conservative: 44
Best Local Similarity: 20.81% Mismatches: 101
Query Match: 8.13% Indels: 92
DB: 5 Gaps: 11
10719236-1_232-1164 (1-311) x BM117875 (1-778)
QY 20 AspGluPheArgGlyCysLeuArgAspGlyLeuPheArgGlyLeuThrAspGlyLeu 39
Db 777 GAGGAATATATACAAATGCGTTGACAAATGTTGTTTGTACCTCGTCAATATCGAATA 718
QY 40 ThrAspThrGluLeuLysSerAlaLysAspIleValIleAspPhePheGluHisGlySer 59
Db 717 CACAGCACTTGGTGTGATGACGTAATTAATTAAGTTCCTTCACAAA--CTT 661
QY 60 GluIleGluLysAspAlaValThrSerProValProThrMetArgGlyPheThrGly 79
Db 660 GAAATGAGAAATTAATTAAGTTCGTGAAATGTCACAAATTC-----GGATACGACTCT 607
QY 80 LeuGluSerGluSerThrAlaGluIleThrAsnThrGlySerTyrosAspTyrosMet 99
Db 606 ATTGAAAGGAAAG-----TTAGATGATGGAAGCCAGGATTAATTAAGCTT 559
QY 100 CysTyrosSerMetGlyThr-----AlaAspAsnLeu 109
Db 558 TCATTTAATACATGCTTGTGCGACGTATTAACGACCATTAATGACCGAGTAATTA 499
QY 110 PheProSerGlyAspPheGluArgIleTythrGlnTyrosPheAspArgGlnTyrosAla 129
Db 498 GTTCCA-----CAATTTCTTCC 481
QY 130 SerArgAlaValAlaArgGluValLeuArgAlaThrGlyThrGluProAspGlyGlyVal 149
Db 480 ACC-----GTG 475
QY 150 GluAlaPheLeu---AspGlyGluProLeu----- 158
Db 474 AAATCATTTTAAAGAAATGCAAAATGCTTGTCTCTCAGAAATTTGATCTTATCATTT 415

QY 158 ----- 158
Db 414 GGGCTAAATTTGACAGACAACTATTCAGCACTTCATGATTCATGATTCATGCA 355
QY 159 -----LeuArgPheArgTyrosPheProGluValProGluHisArgSerAla 173
Db 354 AATATCATTTCTTCGTTACGGTGCCTTATATCA-----CTTATTACAGAAAGTTA 301
QY 174 GluGluGlnProLeuArgMetAlaProHisTyrosLeuSerMetValThrLeuIleGln 193
Db 300 GAGAAAGTTCAAATATGATGCTTGAACACACCGATTATGCGCTTAATGACACTTCTCTT 241
QY 194 GlnThrProCysAlaAsnGlyPheValSerLeuGlnAlaGluValGlyAlaPheThr 213
Db 240 CAAGAT-----TCTGTGTGATTCGATTCAGATTAATTCGCA--CATGGACTTATATC 190
QY 214 AspLeuProTyrosProAspAlaValLeuValPheCysGlyAlaIleAlaThrLeuVal 233
Db 189 GATGCGTCCCAATACAGACACTGTTTGATATAA-TGTGGCAGATTTGTTACAAATACG 131
QY 234 -ThrGlyGlnValAlaLysAlaProArgHisHisValAlaAlaProArgArgAspGln 253
Db 130 GACACATGGAGACCTAAATCTACGCCACATCGCATCTTGTTCAGATCAACTAA 71
QY 253 eAlaGlySerSerArgThrSerSerValPhePheLeuArgProAsnAlaAsp 270
Db 70 GAGGCGTACCCACGACAAATGATGCTTCTTTGTTCATCTTCTGAGGAC 19

Search completed: September 19, 2005, 17:59:29
Job time : 3701 secs


```
Db 232 ATGACACGACGCTGCTCCACCTTCAGCTGCGCCGACCTCCAGCAGGCGCTCCACCGAC 291
Qy 21 GluPheArgArgCysLeuArgAspLysGlyLeuPheTyrLeuThrAspCysGlyLeuThr 40
Db 292 GAGTCCCGCAGGTGCTCTGAGGAGCAAGGGCCCTCTTCTATCTGACGAGCTGGCGTCTGACC 351
Qy 41 AspThrGluLeuLysSerAlaLysAspLysLeuValIleAspPhePheGluHisGlySerGlu 60
Db 352 GACACCGAGCTGAATCGGACCGCAAGGACATCGTCATGCACTTCTTCGAGCAGCGAGCGAG 411
Qy 61 AlaGluysArgAlaValThrSerProValProThrMetCysArgGlyPheThrGlyLeu 80
Db 412 GCGGAGAGCGCGCCCTCACTCCGCCCTCCCAACCACTCCCGCGGCTTCACCGCGGCTG 471
Qy 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys 100
Db 472 GAGTCGAGAGACACCGCCCAATCACCAATACCGGCACTACTCCGACTACTCCATGTC 531
Qy 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleThrThr 120
Db 532 TACTGATGGGACCGCGGACAACTCTCCGTCGGGTGACTTCGAGCGGATCTGAGACC 591
Qy 121 GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
Db 592 CAGTACTTCGACCGCGCACTACACCGCTCCCGCGGCTCCCGGAGGCTCTCGGCGG 651
Qy 141 ThrGlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArg 160
Db 652 ACCGGGACCGAGCCCGAGCGGGGTCCAGAGGCTTCTTCGACTGGACCCCTGCTCGG 711
Qy 161 PheArgTyrPheProGlnValProGluHisArgSerAlaGluGluGlnProLeuArgMet 180
Db 712 TTCGCTACTCTCCGAGGATCCCGAGGACCGGACCGCGAGAGACAGCCCTCGGAGTG 771
Qy 181 AlaProHisTyrAspLeuSerMetValThrLeuIleGlnIleThrProCysAlaAsnGly 200
Db 772 GCGCGGCACTAGACCTTCGATGCTCACTCCATCCAGCAGACACCTGGCGCAAGCGC 831
Qy 201 PheValSerLeuGlnAlaGluValGlyValAlaPheThrAspLeuProTyrArgProAsp 220
Db 832 TTCGTCACCTTCAGCGCGAGGTCCGCGCGGCTTCACGAGACTGCTTACCTCCGTCG 891
Qy 221 AlaValLeuValPheCysGlyValAlaIleAlaThrLeuValThrGlyGluValLysAla 240
Db 892 GCCGCTCTGCTTCTTCGCGGCGCATCCGACCTGGTGAACCGCGCGCAAGTCAAGGCC 951
Qy 241 ProArgHisIleValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260
Db 952 CCCCGGACCATGTCGCGGCGCCCGCGAGGAGCAAGATAGGGGAGCAGCGGACCTCC 1011
Qy 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280
Db 1012 AGTGTGTTCTTCTTCGCGCCCAACCGGACCTTCTCCGCTCCCTGCGCGGCGCAG 1071
Qy 281 CysGlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTyrIleGlyGly 300
Db 1072 TCGCGCTTCGATGTAGCTGAGCGGAGCAGACCGGCACTTCCAGATTGATCGGGGGC 1131
Qy 301 AsnTyrValAsnIleArgArgThrSerLysAla 311
Db 1132 AACACGAGACATCCGCGGACATCCAAAGGCA 1164

RESULT 2
US-08-592-411-14
; Sequence 14, Application US/08592411
; Patent No. 5726032
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Process for the Efficient Production of
; TITLE OF INVENTION: 7-ADCA via 2-(Carboxymethylthio)acetyl-7-ADCA and
; TITLE OF INVENTION: 3-(Carboxymethylthio)propionyl-7-ADCA
; NUMBER OF SEQUENCES: 17
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,411
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 939 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEetical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: No. 5726032ardia lactamdurans
INDIVIDUAL ISOLATE: ATCC 27382
FEATURE:
NAME/KEY: CDS
LOCATION: 1..939
OTHER INFORMATION: /gene= "cefE"
US-08-592-411-14

Alignment Scores:
Pred. No.: 1,99e-142 Length: 939
Score: 1210.00 Matches: 220
Percent Similarity: 83.87% Conservative: 40
Best Local Similarity: 70.97% Mismatches: 50
Query Match: 73.96% Indels: 0
Dbs: 1 Gaps: 0

10719236-1_232-1164 (1-311) x US-08-592-411-14 (1-939)
Qy 2 AspThrThrAlaProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAspGlu 21
Db 7 GACCGGACGCTGCGGACCTTCGATCTGCGGAGCTGCTGAGGCTTTCACCGAGAGAG 66
Qy 22 PheArgArgCysLeuArgAspLysGlyLeuPheTyrLeuThrAspCysGlyLeuThrAsp 41
Db 67 TTCGCGCAGTCTGCGCGGAGGAGGCGGTGTCTTACTTCAAGGACACCGGGCTCGCGAG 126
Qy 42 ThrGluLeuLysSerAlaLysAspLysLeuValIleAspPhePheGluHisGlySerGluAla 61
Db 127 GCGGACCAACCGCTCCGCGCGGAGATCGCGGTGACTTCTTCACACCGGACCGAGGCC 186
Qy 62 GluysArgAlaValThrSerProValProThrMetCysArgGlyPheThrGlyLeuGlu 81
Db 187 GAGAGAGAGCGGTGAGAGACCGCATCCGACCATCCGCGCGGATACCGCGGGCTGGAG 246
Qy 82 SerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCysTyr 101
Db 247 TCGAGAGGACCGCGGAGATCAGAACACCGGCAAGTACACCGCACTACTGATGTCGTAC 306
Qy 102 SerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleThrPheGln 121
Db 307 TCGATGGGACCGCGGACCACTGTTCCAGGCGCGGAGTTCAGAGAGGCGTGGAGGAC 366
Qy 122 TyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAlaThr 141
Db 367 TACTTCGCGGAGATGACCGCGCTTCGACAGAGCTCCGCGGAGAGTCTGACTCGCTC 426
Qy 142 GlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArgPhe 161
Db 427 GCGCGGAGACCGGAGGTGCGGATGAGAGCGCTTCTCGACTGCGGACCCCTGCTGCGCTG 486
Qy 162 ArgTyrPheProGlnValProGluHisArgSerAlaGluGluGlnProLeuArgMetAla 181
Db 487 CGTACTTCTCCCGAGGTGCCCGAGGATCGGTGGCCGAGAGGACCGCGCTGCGGATGGCC 546
```

```

OY      182 Prohi$tyr$apleusereetVal1ThrLeuIlegInntPrCoCYAAlAsnGlyPhe 201
Db      547 CCGCAGTACCACTCTTCGATCGTCAACCCTGATCCACAGACCCCCTTGCGCGAACGGGTTC 606
OY      202 ValSerleuGlnalagluValgIyglValapherThraapleProTyArzProaspAla 221
Db      607 GTGAGCCTTGACGGGTGCAGGGGTGCAGGGGTCTTATGTGGACATCTCCGCCACCCGGGGCGG 666
OY      222 ValIeuValPheCy$glValalalea1athrLeuVal1ThnglyGlnValIeyalaPro 241
Db      667 GTGCTGTGTCTTCCGGCGCGCGGGGAGCGACTGGTGGCCGACGAGGGCGCATTAAGCGGCC 726
OY      242 ArgHisHisValAlaalaaProArgArGaapGlnlaleaglYseSerArzThSerSer 261
Db      727 AAGCACCAAGCTGGCCGGCGCCGGCGCGAGCAAGCGGGTGGGAGAGCGCGCACTCCAGC 786
OY      262 ValPhePheleuarProasnaLaapPheTrPheSerValProleualarGluCys 281
Db      787 GTGTCTTCTCGTGGCCCCCAACGGGGACTTCCGCTCTCGGTGCCGGGGCCAGGGAGTGC 846
OY      282 GlYPhea$paValSerleuaspglyGuThrralathrPhagiNa$PTripileGlyYasn 301
Db      847 GGGTTGACGTCAGCATCTCCGGCCGAGACCGCCACCTTCGACGACTGATCGCGCGCAC 906
OY      302 TyrValasnlleaargatThserIySala 311
Db      907 TACATCAACATCCGAAGAAGACGGCGCGCC 936

RESULT 3
US-08-591-501-15
; Sequence 15, Application US/08591501
; Patent No. 5795733
; GENERAL INFORMATION:
; APPLICANT: BOVENBERG, ROELOF ARY IANS
; APPLICANT: KOEKMAN, BERTUS PIETER
; APPLICANT: HOEKEMA, ANDREAS
; APPLICANT: VAN DER LAAN, JAN METSKE
; APPLICANT: VERWEIJ, JAN
; TITLE OF INVENTION: PROCESS FOR THE EFFICIENT PRODUCTION OF
; TITLE OF INVENTION: 7-ADCA VIA 3-(CARBOXYETHYLTHIO) PROBIONYL-7-ADCA
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVENUE, NW
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591,501
; FILING DATE: 13-MAY-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 24615--20065.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 867-1500
; TELEFAX: (202) 822-0168
; TELEX: 90-4030 MRSNFORSMWSH
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 939 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-591-501-15
```

Alignment Scores:			
Pred. No.:	1,99e-142	Length:	939
Score:	1210.00	Matches:	220
Percent Similarity:	83.87%	Conservative:	50
Best Local Similarity:	70.97%	Mismatches:	40
Query Match:	73.96%	Indels:	0
DB:	1	Gaps:	0
10719236-1_232-1164 (1-311) x US-08-591-501-15 (1-939)			
QY 2 AsprThrValProThrPheSerLeuAlaGluLeuGlnGlyLeuHisGlnAspGlu 21			
DB 7 GACGCGACCGGTGCGACCTTCGACTGTGGCGGACGCTGGGTGAGGGCTTGCACAGAGAG 66			
QY 22 PheArgArgCysLeuArgAspLysGlyLeuPheTyLeuThrAspCysGlyLeuThrAsp 41			
DB 67 TTCGCGCACCTGCTGGCGGAGAAAGGGCGGTGTTCTACCTCAAGGGGACCGGGCTGGCGAG 126			
QY 42 ThrGluLeuLysSerLysAlaSerLysLysLysLysLysLysLysLysLysLysLysLys 61			
DB 127 GCGGACACACGCTTCGGCGGGGAGAGTGGCGGAGCTTTCGACACAGGACGACGAGGCGC 186			
QY 62 GluLysArgAlaValThrSerProValProThrMetCArgArgGlyPheThrGlyLeuGlu 81			
DB 187 GAGAGAGAGCGGTGATGACGCCATCCGACCATCCGGCGGGTACCGCGGGCTGGAG 246			
QY 82 SerGluSerThrAlaGlnLeuThrAsnThrGlySerTySerAspTySerMetCysTy 101			
DB 247 TCCGAGAGCACCGCGAGATTCAGAACCCGGCAAGTACACCGCATCTGATGTCGTAC 306			
QY 102 SerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgGlyLeuThrGln 121			
DB 307 TCGATGGAGCACCGCGGACCACTGTTCCCGACGCGGAGTTGAGAAAGCGTGGAGAGAC 366			
QY 122 TyrPheAspArgGlnTyThrAlaSerArgAlaValAlaArgGluValLeuArgAlaThr 141			
DB 367 TACTTGGCGCGGATGATGACCGGCTTCGACAGACGTGCGGGGAGAGTGTACTCGGTC 426			
QY 142 GlyThrGluProAspGlyValGluAlaPheLeuAspCysGluProLeuLeuArgPhe 161			
DB 427 GGGCGGGAACCCGAGTGGGATGAGACGCTTCCGACTCGACTCGAAACCCCTGCTCGCTG 486			
QY 162 ArgTyPheProGlnAlaProGlnHisArgSerLysGluGlnGlnProLeuArgMetAla 181			
DB 487 CGCTACTTCCCGAGGTGGCCGAGATGCGGTGGCGGAGGACAGCGCTGCGGATGGCC 546			
QY 182 ProHisLysTyPheLeuSerMetValThrLeuLysGlnGlnThrProCysAlaAsnGlyPhe 201			
DB 547 CCGCATGACGACTCTCGATGCTGACCCGTATCCACAGACCCCTTGGCGGAACGGGTTCC 606			
QY 202 ValSerLeuGlnAlaGluValGlyValAlaPheThrAspLeuProTyTrArgProAspAla 221			
DB 607 GTCAGGCTGAGGTTCAGGTGAGCGGGTCTCATGTGACATCCCGGCGAGCCGGGCGG 666			
QY 222 ValLeuValPheCysGlyAlaLysLysLysLysLysLysLysLysLysLysLysLysLys 241			
DB 667 GTCTGTGTTCTTCTGCGCGCGGGGAGCGAGCTGTGTGGCCGACGGCGGATCAAGGCCGCC 726			
QY 242 ArgHisHisValAlaAlaProArgArgAspGlnLysLysSerArgTySerSer 261			
DB 727 AACGACACAGTGGCGCGCGCGCGGACAAAGCGGAGTGGGACACACCGGACTCCAGC 786			
QY 262 ValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCys 281			
DB 787 GTGTCTTCTCTGGCGCCCAAGGGAGCTTCGCTTCTCGGTCGCGCGGACAGGAGTGC 846			
QY 282 GlyPheAspValSerLeuAspGlyGlnThrAlaThrPheGlnAspTrpLysGlyValAsn 301			
DB 847 GGGTTCAGCTCAGCATCCCGGCGAGACCGGCACTTTCAGACGACTGGATCGGCGGAC 906			
QY 302 TyrValAsnLysArgArgThrSerLysAla 311			

```
Db      907  TACATCAACATCCGAGAACCGCGCGCC 936

RESULT 4
US-08-592-411-16
; Sequence 16, Application US/08592411
; Patent No. 5726032
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Process for the Efficient Production of
; TITLE OF INVENTION: 7-ADCA via 2-(Carboxymethylthio)acetyl-7-ADCA and
; TITLE OF INVENTION: 3-(Carboxymethylthio)propionyl-7-ADCA
; NUMBER OF SEQUENCES: 17
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,411
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 942 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: No. 5726032ardia lactamdurans
; INDIVIDUAL ISOLATE: LC 411
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..942
; OTHER INFORMATION: /gene="cefE"
; OTHER INFORMATION: /citation= ([1])
; PUBLICATION INFORMATION:
; AUTHORS: Coque, J J
; AUTHORS: Martin, J F
; AUTHORS: Liras, P
; TITLE: Characterization and expression in
; TITLE: Streptomyces lividans of cefD and cefE genes from
; TITLE: No. 5726032ardia lactamdurans: the organization of the
; TITLE: cephamycin gene cluster differs from that in
; TITLE: Streptomyces clavuligerus
; JOURNAL: Mol. Gen. Genet.
; VOLUME: 236
; PAGES: 453-458
; DATE: 1993
; US-08-592-411-16

Alignment Scores:
Pred. No.: 1,35e-140 Length: 942
Score: 1195.50 Matches: 219
Percent Similarity: 83.28% Conservative: 40
Best Local Similarity: 70.42% Mismatches: 51
Query Match: 73.07% Indels: 1
DB: 1 Gaps: 1

10719236-1_232-1164 (1-311) x US-08-592-411-16 (1-942)

Qy      2  AspThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAspGlu 21
       |||||
Db      7  GACGCGACCGTGCACCTTCGATCTGGCCGCGCTGAGCGCTTCACACGAGGAG 66
Qy     22  PheArgAGCyGLeuArgAspGlyGlyLeuPheTyrLeuThrAspCyGlyLeu---Thr 40
       |||||
Db     67  TTCGCGCACTGCTCGCGAGAGGCGCGTTCACCTCAAGGACACCGCGCTGCGCGCC 126
Qy     41  AspThrGluLeuSerAlaLysAspIleValIleAspPhePheGlnHisGlySerGlu 60
       |||||

Db      127  GAGCGGAGCACGCGCTCGGGCGGGAGATCGGGTGAGACTTCTTCGACCAACGACCGAG 186
Qy      61  AlaGluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeu 80
       |||||
Db     187  GCCGAGAGAGAGCGGTGATGAGCCGATCCGACATCCGCGCGGGTACCGCGGAGCTG 246
Qy      81  GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys 100
       |||||
Db     247  GAGTCGAGAGCACCGCGCATACGACACCGGCAAGTACACCACTACTGATGATG 306
Qy     101  TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGlnArgIleTyrThr 120
       |||||
Db     307  TACTCGATGGGACCGCGGACCACTGTTCCGACGCGGAGAGTTCAGAGAGCGGTGGAG 366
Qy     121  GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuAGla 140
       |||||
Db     367  GACTACTCCGGGAGTATCCCGGCTTGACAGACGTGCGGGGAGAGGTGACTGACCTG 426
Qy     141  ThrGlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArg 160
       |||||
Db     427  GTGCGCGCGGACCCGAGGTGCGCATGAGACGCTTCTCGACTGCGAACCCCTGCTGCG 486
Qy     161  PheArgTyrPheProGlnValProGlnHisArgSerAlaGluGlnProLeuArgMet 180
       |||||
Db     487  CTGGGCTACTTCCCGAGGTGCCGAGGATCGCGTGGCGGAGGACGCGCTGCGGATG 546
Qy     181  AlaProHisTyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGly 200
       |||||
Db     547  GCCCGGACATACACCTCTGATCGTACCCCTGATCACAGACCCCTTGGCGGAGCGG 606
Qy     201  PheValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrArgProAsp 220
       |||||
Db     607  TTCTGACGCTGACAGGTGACAGGTGACGGGTCTATGTGACATCCCGCGGACCGGGG 666
Qy     221  AlaValLeuValPheCysGlyAlaIleAlaThrLeuValThrGlyGlnValVala 240
       |||||
Db     667  GCGGTGCTGTGTTTGGCGGCGGTGGCGACCTGCTGCGCCGACCGGATCAAGGGG 726
Qy     241  ProArgHisHisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260
       |||||
Db     727  CCAAGACACCACTGGCGCGCGCGGCGGACAAACGGGTGGACAGACCGCACCTCC 786
Qy     261  SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgIle 280
       |||||
Db     787  AGCGTGTCTTCTGCGCCCAACCGGGAATTCGCTCTCGGTGCGGGGCAAGAG 846
Qy     281  CysGlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrpIleGlyGly 300
       |||||
Db     847  TGGCGGTTGACGTCAGCATCCCGCGCGAGACCGCACCTTCAGACGATGATGCGCGG 906
Qy     301  AsnTyrValAsnIleArgArgThrSerIysAla 311
Db     907  AACTACATCAACATCCGAGAACCGCGCGCC 936

RESULT 5
US-08-591-501-14
; Sequence 14, Application US/08591501
; Patent No. 5795733
; GENERAL INFORMATION:
; APPLICANT: BOVENBERG, ROELOF ARY LANS
; APPLICANT: KOERMAN, BERTUS PIETER
; APPLICANT: HOEKEMA, ANDREAS
; APPLICANT: VAN DER LAAN, JAN METSKE
; APPLICANT: VERMEIJ, JAN
; APPLICANT: DE VROOM, ERIK
; TITLE OF INVENTION: PROCESS FOR THE EFFICIENT PRODUCTION OF
; TITLE OF INVENTION: 7-ADCA VIA 3-(CARBOXYETHYLTHIO) PROPIONYL-7-ADCA
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVENUE, NW
; CITY: WASHINGTON
; STATE: DC
```

COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/591,501
FILING DATE: 13-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ADLER, REID G.
REGISTRATION NUMBER: 30,988
REFERENCE/DOCKET NUMBER: 24615-20065.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 822-0168
TELEX: 90-4030 MRSNFORSMH
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 942 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
PUBLICATION INFORMATION:
AUTHORS: COQUE et al.,
JOURNAL: Mol. Gen. Genet.
VOLUME: 236
PAGES: 453-458
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 14: FROM 1 TO 942
US-08-591-501-14

Alignment Scores:
Pred. No.: 1,35e-140 Length: 942
Score: 1195.50 Matches: 219
Percent Similarity: 83.28% Conservative: 40
Best Local Similarity: 70.42% Mismatches: 51
Query Match: 73.07% Indels: 1
DB: 1 Gaps: 1

10719236-1_232-1164 (1-311) x US-08-591-501-14 (1-942)
QY 2 AAPTThThValProThrPhSeSerLeuAlaGluLeuGlnGlyLeuHISgluAspGlu 21
DB 7 GACGGACCGCTGCCGACCTTCATCTGGCCGAGCTGCGGAGGGCTTGCACCGAGAGGAG 66
QY 22 PheArgArgCysLeuArgAspLysGlyLeuPheTyrLeuThrAspCysGlyLeu---Thr 40
DB 67 TTCCCGCACTGCTCTGCGCGAGAGAGGGCGGTTCCTACTCAAGGGCACCGGGCTGCCGGCC 126
QY 41 AAPTThGluLeuLysSerAlaLysAspIleValIleAspPheGluHISglYserGlu 60
DB 127 GAGGGGACCAACGCTCGGGCGGAGATCGCGTGAATCTTTCGACACGCGGACCGAG 186
QY 61 AlaGluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeu 80
DB 187 GCCGGAAGAAAGCGGTGTATGACCGGATCCGACCATCCGGCGGGGATGCCCGGCTG 246
QY 81 GluSerGluSerThrAlaGlnIleThrAnThrGlySerTyrSerAspTyrSerMetCys 100
DB 247 GAGTCGAGAGACACCGCGGAGATCAAGAACCGCGCAACTACCACTACTCGATGCG 306
QY 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTyrThr 120
DB 307 TACTCGATGGGACCGCGGACAACTGTTCGCCACGCGGAGTTGAGAAAGCGGTGGAG 366
QY 121 GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
DB 367 GACTCTCTTGGCGGGATGTACCGGCTTGCAGGAGCGTCCCGGCGAGGTGTGACCTCG 426
QY 141 ThrGlyThrGluProAspGlyGlyValAlaGluAlaPheLeuAspCysGluProLeuLeuArg 160

DB 427 GTCCGCGGGAACCGGAGAGTGGCATGACGCTTCTCCGACTGGAACCCCTGCGCC 486
QY 161 PheArgTyrPheProGlnValProGluHISargSerAlaGluGlnProLeuArgMet 180
DB 487 CTGGCTACTTCTCCCGAGAGTGCCCGAGGATGCGCGTGGCGAGAGCACCCCTGCGGATG 546
QY 181 AlaProHISArgPheLeuSerMetValThrIleuIleGlnGlnThrProCysAlaAsnGly 200
DB 547 GCCCGCACTAGACCTTCATCGATCACTGATCCACCGAGACCCCTTGGCGGAACGGG 606
QY 201 PheValSerLeuGlnAlaGluValGlyValaPheThrAspLeuProTyrArgProAsp 220
DB 607 TTCCGACACCTCGAGAGTGGAGTGGAGCGGTCTTATGTGACATCCCGCGGACCGGCGC 666
QY 221 AlaValLeuValaPheCysGlyAlaIleAlaThrLeuValaThrGlyGlyValaIleAla 240
DB 667 GCGGTGCTGTGTCTCTGCGCGCGGTGGCGACGCTGTGGCGGACGGCGCATCAAGGCG 726
QY 241 ProArgHISHisValaIleAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260
DB 727 CCAGACACCAAGTGGCGCGCGCGCGCGGACCAAGCGGAGGCGAGCGGACCACTCC 786
QY 261 SerValPhePheLeuArgProAsnAlaAspPheThrPhSeSerValProLeuAlaArgGlu 280
DB 787 AGCGTCTTCTCTCGCGCCCAACGGGACCTTCCGCTTCTGCTGCCGGGCGGAGGAG 846
QY 281 CysGlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTyrIleGly 300
DB 847 TGCGGTTGACAGTCAGATCCCGCGGAGACCGGACCTTGCAGCACTGATCGGCGGC 906
QY 301 AsnTyrValaIleuIleArgArgThrSerVala 311
DB 907 AACTACATCAACATCCGGAAGACCGCGCGCC 939

RESULT 6
PCT-US95-04801-2
Sequence 2, Application PC/TUS9504801
GENERAL INFORMATION:
APPLICANT: Martin, Juan F.
APPLICANT: Coque, Juan R.
APPLICANT: Enguita, Francisco J.
APPLICANT: Fuente, Juan L.
APPLICANT: Ilarena, Francisco J.
APPLICANT: Liras, Paloma
TITLE OF INVENTION: DNA ENCODING CEPHARYCIN BIOSYNTHESIS
TITLE OF INVENTION: LATE GENES
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: John W. Wallen III
STREET: P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04801
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Wallen III, John W.
REGISTRATION NUMBER: 35,403
REFERENCE/DOCKET NUMBER: 19179
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3905
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:


```

; LENGTH: 972 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US95-04801-2

```

```

Alignment Scores:
Pred. No.: 7.58e-112 Length: 972
Score: 967.50 Matches: 186
Percent Similarity: 73.84% Conservative: 37
Best Local Similarity: 61.59% Mismatches: 78
Query Match: 59.14% Indels: 1
DB: 5 Gaps: 1
10719236-1_232-1164 (1-311) x PCT-US95-04801-2 (1-972)

```

```

Oy 4 ThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAspGluPheArg 23
Db 4 ACGGTACCGGTCTTCAGCATGGCCCAACTGGCCGACGGCTGCGCCGACGAGATTCGCC 63
Oy 24 ArgCysLeuArgAspGlyLeuPheTyrLeuThrAspCysGlyLeuThrAspGlu 43
Db 64 GAGTGGGCCCGC---CGCGGGGTCTTCTACCTCACCGGGGTACGGGCGCACCGAAGAGAC 120
Oy 44 LeuysSerAlaIysAspIleValIleAspPhePheGlnHisGlySerGluAlaGluIys 63
Db 121 CACCGGTGGCCACCGACACCGCGATGACTTCTTCCGCCACGACCGCCGACGAGAGAG 180
Oy 64 ArgAlaValThrSerProValProThrMetArgArgGlyPheThrGlyLeuGluSerGlu 83
Db 181 CAGGCGGTGACCAAGATGCTCCGACCATGGCGCGGGGTACTCGGGCTGAGGCGGAA 240
Oy 84 SerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCysTyrSerMet 103
Db 241 AGCACCGCCGAGTCTACCAACACCGCGACCTACCGACTCATGCTGACTCGATG 300
Oy 104 GlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgGlyIleThrGlnTyrPhe 123
Db 301 GGCATCGCGCGACCTTCTCCGTCGAAAGAGTTCGAGTGGTCTGACCGACTACTTC 360
Oy 124 AspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAlaThrGlyThr 143
Db 361 GACAGCTGTACCGCGCGCGACGAGACCGGGCGCTGAGTGTACCGCGCGGCGACC 420
Oy 144 GluProAspGlyGlyValGluAlaPheLeuAspCysGlyProLeuLeuArgPheArgTyr 163
Db 421 TACGACGCGGACGACTCGACACCTGCTGCTGACGACCGGCTGCGCTGCGGTAC 480
Oy 164 PheProGlnValProGlnHisArgSerAlaGluGlnGlnProLeuArgMetAlaProHis 183
Db 481 TTCGCGAGGTCCCGAGACACCGCGCCGACGATGACGACACCGCGGATGCGCGGAC 540
Oy 184 TyrAspLeuSerMetValThrIleuIleGlnGlnThrProCysAlaAsnGlyPheValSer 203
Db 541 TACGACCTGTCCATCATCATTCACACGACCCCGTGGCGCAAGGTTTGTCTGACG 600
Oy 204 LeuGlnAlaGluValGlyValAlaPheThrAspLeuProTyrArgProAspAlaValLeu 223
Db 601 CTGCAAGCGGACGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 660
Oy 224 ValPheCysGlyAlaIleAlaThrIleuValThrGlyGlyGlnValAlaProArgHis 243
Db 661 GTGCTGTGGGCGCATGCGCGCTGTGTCACCCACGAGGCGCGGTCCCGCGCCACACAC 720
Oy 244 HisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSerValPhe 263
Db 721 CACGTGCTCTCCCGGACGCGACGATGCTCAAGGACGACGACGACGACGACGACGACGAC 780
Oy 264 PheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCysGlyPhe 283
Db 781 TTCCTGCGCGCGCGCATTTCACTTCTCGGTGCGCGACGCGACGAGAGATGACGCGCTC 840

```

```

Oy 284 AspValSerLeuAspArgGlyIleThrAlaThrPheGlnAspTrpIleGlyIysAsnTyrVal 303
Db 841 GACGTACGCTTGACATGAGAGAGGACCTTGGGACTGAGTGGAGACCAACTACGTC 900
Oy 304 AsnIle 305
Db 901 ACGATG 906

```

RESULT 7

```

US-08-379-556A-9
; Sequence 9, Application US/08379556A
; Patent No. 5859329
; GENERAL INFORMATION:
; APPLICANT: HOLTON, TIMOTHY A.
; APPLICANT: KEAM, LISA A.
; TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOL
; TITLE OF INVENTION: SYNTHASE ENZYMES AND USES THEREFORE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
; STREET: 400 GARDEN CITY PLAZA
; CITY: GARDEN CITY
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 11530

```

COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/379,556A
FILING DATE: 22-MAR-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: DIGILIO, FRANK S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 9592
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516)742-4343
TELEFAX: (516)742-4366
TELEX: 230 901 SANS UR

```

INFORMATION FOR SEQ ID NO: 9:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 1250 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 11..1015
US-08-379-556A-9

```

Alignment Scores:

```

Pred. No.: 1.49e-10 Length: 1250
Score: 165.00 Matches: 72
Percent Similarity: 40.19% Conservative: 55
Best Local Similarity: 22.78% Mismatches: 135
Query Match: 10.09% Indels: 54
DB: 2 Gaps: 14
10719236-1_232-1164 (1-311) x US-08-379-556A-9 (1-1250)

```

```

Oy 5 ValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAspGluPheArg 24
Db 137 GTTCCAGTATGATTCATTCACCGGCGGATTAAGAACTCTTGCTGCTTGATATCCAA 196
Oy 25 CysLeuArgAspGlyLeuPheTyrLeuThrAspCysGlyLeuThrAspThrGluLeu 44
Db 197 GCGAGCAAGATGGGGTATTTTCAAGTGTGAACCATGGATACCAAGTGAATCATTT 256

```



```

Qy 45 LysSerAlaLysAspIleValIleAspPhePheGluHisGlySerGluIaGluLysArg 64
Db 257 AGCAATTACAAATGTGAAAAAGGTTCTTGACCTCCACAAAGAA---GAAAAAGAA 313
Qy 65 AlaValThrSerProValProThrMetCArgGlyPheThrGlyLeuGluSerGluSer 84
Db 314 GTCATTGTCAAAACCA-----GATGGTTACAAAGGTCGTTGAA----- 349
Qy 85 ThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCysTyrSerMetGly 104
Db 350 -----GGGTATGGAACAAAGCTTCAAAAAGGCTTGACGAAGCAAGAAAGCA 394
Qy 105 ThrAlaAspAsnLeuPhe-----ProSerGlyAspPheGluArgIleTyr 119
Db 395 TGGGTGATCATTTGTTTCATATAGTTTGCCACCTTCTTTATTAATCATCATATTG 454
Qy 120 ThrGln-----TyrPheAsp-----ArgGlnTyrThrAlaSerArg---Ala 132
Db 455 CCAAAGAACCCCTCCTTACAGAGATACAAATGAGAGATACACAAAGCTTGATAGGG 514
Qy 133 ValAlaArgGluValLeu-----ArgAlaThrGlyThrGluProAsp----- 146
Db 515 GTGGCAATTAAGTTCCTGGCTGTGTGCAAAAAGGCTTGAGTGAAGAATGAAGTGC 574
Qy 147 -----GlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArgPheArg 162
Db 575 AAGCAAGCCTGGGGGCGAAGCCTGTGATCTAC-----ATGTTGAAATTAAC 622
Qy 163 TyrPheProGlnValProGluHisArgSerAlaGluGluGlnProLeuArgMetAlaPro 182
Db 623 TACTACCCACCATGTCCA-----TGCCCGGACCTGCTCTTGCGGATAGCCCA 670
Qy 183 HisTyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaHisGlyPheVal 202
Db 671 CATACCGACATGTCTTCAATCAACATCACTT-----GTCCCGAATTAAGTTCAA 718
Qy 203 SerLeuGlnAlaGluValGlyAlaPheThrAspLeuProTyrArgProAspAlaVal 222
Db 719 GGTCTACAAAGTCTTTAAAGATGTCATGATGATGATGCTTACATTCCTATGCTCTC 778
Qy 223 LeuValPheCysGlyValAlaIleAlaThrLeuValThrGlyGlyValIleValAspArg 242
Db 779 ATTATTACATTTGGTACACAGATGATGATGATGATGATGATGATGATGATGATGAT 838
Qy 243 HisHisValAlaAlaPheArgArgArgArgArgArgArgArgArgArgArgArgVal 262
Db 839 CACAGATCAACTGTGATTAAGAG-----AAAACAAGATGTCGTGGCCA 883
Qy 263 PhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCysGly 282
Db 884 GCATTTTGTGAGCCACCGCCAGAGTTTGAGTTGCT--CCAATTCCAAAAGCTCGTCAT 940
Qy 283 PheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTyrPle 298
Db 941 AAAAGCATCCACCAAAATACAAAGCAAGAAGTACAAAGACTATGTC 988

```

```

; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: oligonucleotide 12
US-08-592-411-12

Alignment Scores:
Pred. No.: 7,01e-11 Length: 105
Score: 154.00 Matches: 29
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 9.41% Gaps: 0
DB: 1

10719236-1_232-1164 (1-311) x US-08-592-411-12 (1-105)

Qy 283 PheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTyrPleGlyValSerTyr 302
Db 1 TTGCATGTGCACCTGTGAGACGGGAGACCGCCAGTTCCAGATGATCGGGGGCAACTAC 60

Qy 303 ValAsnIleArgArgThrSerIleAla 311
Db 61 GTGAACATCCGCCGACATCCAGGCA 87

RESULT 9
US-08-591-501-12
; Sequence 12, Application US/08591501
; Patent No. 579573
; GENERAL INFORMATION:
; APPLICANT: BOVENBERG, ROELOF ARY LANS
; APPLICANT: KOEKMAN, BERTUS PIETER
; APPLICANT: HOEKEMA, ANDREAS
; APPLICANT: VAN DER LAAN, JAN METSKE
; APPLICANT: VERWEIJ, JAN
; APPLICANT: DE VROOM, ERIK
; TITLE OF INVENTION: PROCESS FOR THE EFFICIENT PRODUCTION OF
; TITLE OF INVENTION: 7-ADCA VIA 3-(CARBOXYETHYLTHIO) PROPIONYL-7-ADCA
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVENUE, NW
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591,501
; FILING DATE: 13-MAY-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 24615-20065.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 822-0168
; TELEX: 90-4030 MRSNFOERSMWSH
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs

```

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-591-501-12

Alignment Scores:

Pred. No.:	7.01e-11	Length:	105
Score:	154.00	Matches:	29
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	9.41%	Indels:	0
DB:	1	Gaps:	0

10719236-1_232-1164 (1-311) x US-08-591-501-12 (1-105)

Qy 283 PheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrpIleGlyGlyAsnTyr 302
Db 1 TTCCATGTCACGCTCGAGCGGAGACCCGCCAGTTCAGATTGGATCGGGGCAACTAC 60

Qy 303 ValAsnIleArgArgThrSerIleAla 311
Db 61 GTGAACATCGCGCCACATCCAGGCA 87

RESULT 10

US-08-379-556A-7

Sequence 7, Application US/08379556A

Patent No. 5859329

GENERAL INFORMATION:

APPLICANT: HOLTON, TIMOTHY A.

APPLICANT: KEAM, LISA A.

TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOL

TITLE OF INVENTION: SYNTHASE ENZYMES AND USES THEREFORE

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS: 21

ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER

STREET: 400 GARDEN CITY PLAZA

CITY: GARDEN CITY

STATE: NEW YORK

COUNTRY: USA

ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/379,556A

FILING DATE: 22-MAR-1995

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: DIGICLIO, FRANK S.

REGISTRATION NUMBER: 31,346

REFERENCE/DOCKET NUMBER: 9592

TELECOMMUNICATION INFORMATION:

TELEPHONE: (516)742-4343

TELEX: 230 901 SANS UR

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1236 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 70..1068

US-08-379-556A-7

Alignment Scores:

Pred. No.:

Score:

4.75e-09

Length:

1236

Matches:

153.00

71

Percent Similarity:	39.02%	Conservative:	57
Best Local Similarity:	21.65%	Mismatches:	126
Query Match:	9.35%	Indels:	74
DB:	2	Gaps:	15

10719236-1_232-1164 (1-311) x US-08-379-556A-7 (1-1236)

Qy 1 MetAspThrThr-----ValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHis 18
Db 184 CTCGACACGGTTCTCGAGGTCCCTCGCAGTCACTCAAGC-----CTCGAG 228

Qy 19 GlnAspGluPheArgArgCysLeu-----ArgAspIysGlyLeuPheTyrLeu 34
Db 229 GAGAGCAGCTGTGAAAACTCGTCTTAGTCAGCAAGCAAGAGCGGGACTGTTCAAGTTC 288

Qy 35 ThrAspCysGlyLeuThrAspThrGluLeuLysSerAlaLysAspIleValIleAspPhe 54
Db 289 ACCAACACCGAATTCGAGCTGAGTCAATGAAATAATTCGAAATAAGTCGGTAAGATGTTT 348

Qy 55 PheGluHisGlySerGluAlaGluLysArgAlaValThrSerProValProThrMetArg 74
Db 349 TTTCGAGCTCCCGCAGAG--GAGAAAGAGACGATTCGAAACCCGAGGCTGTGTGAA 405

Qy 75 ArgGlyPheThrGlyLeuGlnSerGlnSerThrAlaGlnIleThrAsnThrGlySerTyr 94
Db 406 GGGTATCGGACCATGTTGCCAAAGAGATTCAGGGGAGA-----444

Qy 95 SerAspTyrSerMetCysTyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAsp 114
Db 445 -----AAAGTTGGGTTGATCATTTG-----465

Qy 115 PheGluArgIleThrPhe-----GlnTyrPhe-----123
Db 466 TTTCACAGGTTGGCTCTAGTGTATTAACTACCAATGCGCTTAAGACTCTTCT 525

Qy 124 -----AspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeu--- 138
Db 526 TATAGGAGGCGCAACGAGAGTACACAAAGTACTTAAGATAGTACCGCAAGCTCTTC 585

Qy 139 -----ArgAlaThrGlyThrGluProAspGly-----GlyVal 149
Db 586 AAGGTATGTCGCAAGGACTTGGTTTGAAGAAAGTAAAGTCAAAAATCATGTGGCAAT 645

Qy 150 GluAlaPheLeuAspCysGluProLeuLeuArgPheArgTyrPheProGlnValProGlu 169
Db 646 GAA-----GACATAGTCTACTTCTCAAGATCACTACTACCCGCTTCTCCGA 696

Qy 170 HisArgSerAlaGluGlnGluProLeuArgMetAlaProHisTyrAspLeuSerMetVal 189
Db 697 -----CCGACCTTGGCTTACAGGGTGGCGGCTCAGACTGACTTGAAGCGTCATC 744

Qy 190 ThrLeuIleGlnGlnThrProCysAlaAsnGlyPheValSerLeuGlnAlaGluValGly 209
Db 745 ACCATCTCT-----GTTCCGAATGATGTTGCCGCTTTCACGCTCTTAGAGAC 792

Qy 210 GlnAlaPheThrAspLeuProTyrArgProAspAlaValLeuValPheCysGlyAlaIle 229
Db 793 GGAAGTTGGTACATGTCATGCTATCTTAATGACTCATCTCAACGCTTGTGACCA 852

Qy 230 AlaThrLeuValThrGlyGlnValValAlaProArgHisIleValAlaAlaProArg 249
Db 853 ATGAGATGATAGCAATGAGTAGAGTACAGCGCGTCTCATATGTCAGCAGTGAACAA 912

Qy 250 ArgAspGlnIleAlaGlySerSerArgThrSerSerValPhePheLeuArgProAsnAla 269
Db 913 GAA-----AGAACAGGATATCTGTGGCCCGCTGCTCGAAGCCGATCA 957

Qy 270 AspPheThrPheSerValProLeuAlaArgGluCysGlyPheAspValSerLeuAspGly 289
Db 958 GACTTTGCAAGTCGGG---CCTATTCCAAAGCTCATTAAGTATGATAAAAGCCAGCAAGTAT 1014

Qy 290 GluThrAlaThrPheGlnAspTrp 297
Db 290 -----

Db 1015 AAGACGAGGTGTTTCCGAGTAC 1038

RESULT 11

US-09-902-540-7970

/ Sequence 7970, Application US/09902540

/ Patent No. 6833447

/ GENERAL INFORMATION:

/ APPLICANT: Goldman, Barry S.

/ APPLICANT: Hinkle, Gregory J.

/ APPLICANT: Slater, Steven C.

/ APPLICANT: Wiegand, Roger C.

/ TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

/ FILE REFERENCE: 38-10(15849)B

/ CURRENT APPLICATION NUMBER: US/09/902,540

/ CURRENT FILING DATE: 2001-07-10

/ PRIOR APPLICATION NUMBER: 60/217,883

/ PRIOR FILING DATE: 2000-07-10

/ NUMBER OF SEQ ID NOS: 16825

/ SEQ ID NO 7970

/ LENGTH: 999

/ TYPE: DNA

/ ORGANISM: Myxococcus xanthus

US-09-902-540-7970

Alignment Scores:

Pred. No.:	9,33e-09	Length:	999
Score:	149.50	Matches:	77
Percent Similarity:	32.49%	Conservative:	39
Best Local Similarity:	21.57%	Mismatches:	118
Query Match:	9.14%	Indels:	123
DB:	4	Gaps:	15

10719236-1_232-1164 (1-311) x US-09-902-540-7970 (1-999)

QY 21 GluPheArgArgCySLeuArgAspLySgLyLeuPheTyrLeuThrAspCySgLyLeuThr 40

Db 70 GAAATCGACGACGCTGCGCGACGACGCGGCTTCTTCTACGTGACGCGGCGACGCGGCTTC 129

QY 41 AspThrGluLeu-----LysSer 46

Db 130 GCGGAGTCTCGCGCGCTGAGCGGAGACCAACCGCTTCTTTCGCGCGCGGCTG 189

QY 47 AlaIAspIleValIle-----AspPheGluHis 57

Db 190 GCGAAGAGGACCATGCCATGTCGATGCGCGGCTGCGCGGAGGCTGGTTTCCACTC 249

QY 58 GlySerGluAlaGluLysArgAlaValIhrSerProValProThrMetArgArgLy--- 76

Db 250 GGGGGGGAG-----CTGACCTCGGGCGCGCCGAGACCGGAGAGGGGCTTG 294

QY 77 PheThrGlyLeuGluSerGluSerThrAlaGlnIleThrAsnThrGly----- 92

Db 295 TATCTGGGACAGGACTCGAAGCGAGCATCCGCGCTAAAGCGCGCTGGCCGTTGCAC 354

QY 93 -----SerTyrSerAspTyr--- 97

Db 355 GCGGCAACCTGTGCGCCCGGAGGTGCGGAGCTGCGGCGCGGCTGCTGCACTACGTG 414

QY 98 -----SerMetCysTyrSerMetGlyThr 105

Db 415 GCGGCTGACGCGCGCGCGCGCATGCGCTGATGAGAGGATGGCGGTGAGCGGCG--- 471

QY 106 AlaAspAsnLeuPheProSerGlyAspPheGluArgIleTrrThrGlnTyrPheAspArg 125

Db 472 -----CTGGACGCGGACTACTTCCGCGCG 495

QY 126 GlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAlaThrGlyThrGluPro 145

Db 496 CACTACACGCGCGGACCCGACGCGT----- 519

QY 146 AspGlyGlyValGluAlaPheLeuAspCySgLyProLeuLeuArgPheArgTyrPhePro 165

Db 520 -----CTCTTCGCGCATCTTCATTACCCC 543

QY 166 GlnValProGluHisArgSerAlaGluGlnProLeuArgMetAlaProHisTyrAsp 185

Db 544 GCGGAGCCCGCAC-----GAGAGGTGACCTGGGCGGTGGGAGACACCGAC 594

QY 186 LeuSerMetValThrLeuIleGlnIhrProCyAlaAsnGlyPheValSerLeuGln 205

Db 595 TACGGGCTGCTACGCTCTGCGCCAGACGAC-----MACGC-----GGTTGCGAG 642

QY 206 AlaGluValGlyGlyAlaPheThrAspLeuProTyrArgProAspAlaValLeuValPhe 225

Db 643 GTGAAGACGCGCGCGGCTGGGTGAGGTACCGCGCTGCCGCGGACCGCTGGTGCAC 702

QY 226 CysGlyAlaIleAlaThrLeuValThrGlyGlyValLysAlaProArgHisIstVal 245

Db 703 ATCGGCAATGCTCGACCGGATGACGCGGCTGTACCGCTCCACCGCGCGCGGTG 762

QY 246 AlaIaProArgArgAspGlnIleAlaGlySerSerArgThrSerSerValPhePheLeu 265

Db 763 -----AAGACGTGACGCGCAAGACAGGCTGTCTTCCACTCTTTC 807

QY 266 ArgProAsnAlaAspPheThrPheSerVal---ProLeuAlaArgGluCysGlyPheAsp 284

Db 808 GACCGG-----GACTTCGCGCGCGAGGTCCACCCCTGCACGCGGCGCGCTGAC 861

QY 285 Val-----SerLeuAspGlyGluThrAlaThr 293

Db 862 GTAGATGATGACCGCGCGCGCTGGATGCGGACGTCACGCTTCAGGGGACG 921

QY 294 PheGlnAspTrrPileGlyIAsnTyrValAsnIleArgArgThrSerLyS 310

Db 922 TACGCGCATCTACCTGCGGCAAGGTTCACAGGTTCCTTAAGAGTCT 972

RESULT 12

US-09-902-540-806

/ Sequence 806, Application US/09902540

/ Patent No. 6833447

/ GENERAL INFORMATION:

/ APPLICANT: Goldman, Barry S.

/ APPLICANT: Hinkle, Gregory J.

/ APPLICANT: Slater, Steven C.

/ APPLICANT: Wiegand, Roger C.

/ TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

/ FILE REFERENCE: 38-10(15849)B

/ CURRENT APPLICATION NUMBER: US/09/902,540

/ CURRENT FILING DATE: 2001-07-10

/ PRIOR APPLICATION NUMBER: 60/217,883

/ PRIOR FILING DATE: 2000-07-10

/ NUMBER OF SEQ ID NOS: 16825

/ SEQ ID NO 806

/ LENGTH: 5051

/ TYPE: DNA

/ ORGANISM: Myxococcus xanthus

US-09-902-540-806

Alignment Scores:

Pred. No.:	1.23e-07	Length:	5051
Score:	149.50	Matches:	77
Percent Similarity:	32.49%	Conservative:	39
Best Local Similarity:	21.57%	Mismatches:	118
Query Match:	9.14%	Indels:	123
DB:	4	Gaps:	15

10719236-1_232-1164 (1-311) x US-09-902-540-806 (1-5051)

QY 21 GluPheArgArgCySLeuArgAspLySgLyLeuPheTyrLeuThrAspCySgLyLeuThr 40

Db 4093 GAAATCGACGACGCGCTGCGGACGAGCGGCTTCTTCTACGTGACGCGGCGGAGGTGCC 4152

QY 41 AspThrGluLeu-----LysSer 46

Db 4153 GCGGAGGTCTCGCGCGCTGAGCGGAGACCGCATCTTCTTGGCGTGCCTCCCGGCTG 4212

```

QY 47 AlAlAspAlIleValIle-----AspPheGluHis 57
Db 4213 GCGAAGAGGCGCATGTCATGCGGCGGTGCGGCGAGCGGTGCTTCACATC 4272
QY 58 GlysGluAlaGluYArgAlaValThrSerProValProThiMetArgIleY--- 76
Db 4273 GGGGGGGAG-----CTGACCTCGGGGGCGCCCGGACCGGAGAGGGGTG 4317
QY 77 PheThrGlyLeuGluSerGluSerThrAlaGlnIleThrAsnThrGly----- 92
Db 4318 TATCTGGGCGACGAGCTCGAAGCGAGCATCCGCGCTGAAGCGCTGCGCTTGAC 4377
QY 93 -----SerTyrSerAspTyr--- 97
Db 4378 GAGCCCAACTGTGCGCCCGGAGGTGCCGAGCTCGCGCGGCGGTGCTGACTGCTG 4437
QY 98 -----SerMetCysTyrSerMetGlyThr 105
Db 4438 GCGGCTGACGCGCGCGCGCATGCGTGAAGGAGCATGGCGCTGAGCTG6GC--- 4494
QY 106 AlaAspAsnLeuPheProSerGlyAspPheGluArgIleTyrThrGlnTyrPheAspArg 125
Db 4495 -----CTGACGCGGACTACTTCCGCGG 4518
QY 126 GlnTyrThrAlaSerArgAlaValAlaArgIleValLeuArgAlaThrGlyThrGluPro 145
Db 4519 CACTACAGCGGCGAGCCGACGCTG----- 4542
QY 146 AspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArgPheArgTyrPhePro 165
Db 4543 -----CTCTCCGATCTTCATACCCC 4566
QY 166 GlnValProGluHisArgSerAlaGluGluGlnProLeuArgMetAlaProHisTyrAsp 185
Db 4567 GCGAGCGCCCGAC-----GAGAGCTGACCTGCGGCGGAGCAGCACCCAC 4617
QY 186 LeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGlyPheValSerLeuGln 205
Db 4618 TACGGGCTGCTCACTGCTGCGCCAGAC-----AACGC-----GGTTGAG 4665
QY 206 AlaGluValGlyGlyAlaPheThrAspLeuProTyrArgProAspAlaValLeuValPhe 225
Db 4666 GTAGAGCGCGCGCGCTGGGTGAGGTACCGCGCTGCCGCGGACCTGCTGTGCAAC 4725
QY 226 CysGlyAlaIleAlaThrLeuValThrGlyGlyGlnValAlaProArgHisHisAla 245
Db 4726 ATCGCGCATGCTCGACCGGATGACAGCGGGGTGTAACGCTCCACCGCCGCGGGGTG 4785
QY 246 AlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSerValPhePheLeu 265
Db 4786 -----AAGAAGCTGAGCGGCAAGACAGGCTCTCTTCCACTTCTTC 4830
QY 266 ArgProAsnAlaAspPheThrPheSerVal---ProLeuAlaArgGlyCysGlyPheAsp 284
Db 4831 GACCCG-----GACTTGGCGCGCAGGTTCACCCCTGCCACGCGCGCGCTGAC 4884
QY 285 Val-----SerLeuAspGlyGluThrAlaThr 293
Db 4885 GTAAGATGATGACCGCGCGCGCTGGATGCGGCGGCGGCTTCACAGGCGCAG 4944
QY 294 PheGlnAspTyrPileGlyGlyAsnTyrValaenIleArgArgThrSerIys 310
Db 4945 TACGGCGACTACTGCTGCGGCAAGGTTCAGAGGTCTCCCTTAGAAGTCT 4995

```

RESULT 13
 US-08-592-411-13
 ; Sequence 13, Application US/08592411
 ; Patent No. 5726032
 ; GENERAL INFORMATION:

; APPLICANT: Process for the Efficient Production of
 ; TITLE OF INVENTION: 7-ADCA via 2-(Carboxyethylthio)acetyl-7-ADCA and
 ; TITLE OF INVENTION: 3-(Carboxymethylthio)propionyl-7-ADCA

NUMBER OF SEQUENCES: 17
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/592,411
 FILING DATE:
 CLASSIFICATION: 435
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 102 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 INDIVIDUAL ISOLATE: oligonucleotide 13

Alignment Scores:
 Pred. No.: 1.22e-09 Length: 102
 Score: 144.00 Matches: 27
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.80% Indels: 0
 DB: 1 Gaps: 0

10719236-1_232-1164 (1-311) x US-08-592-411-13 (1-102)

```

QY 284 AspValSerLeuAspGlyGlyThrAlaThrPheGluAspTyrPileGlyGlyAsnTyrVal 303
Db 1 GATGTGCGCTGAGCGGCGGAGCCGACGTCCTCAGAGATTGATCGGGGCAACTACGTG 60
QY 304 AsnIleArgArgThrSerIys 310
Db 61 AACATCGCGCGCATCCAG 81

```

RESULT 14
 US-08-591-501-13
 ; Sequence 13, Application US/08591501
 ; Patent No. 5795733
 ; GENERAL INFORMATION:

; APPLICANT: BOVENBERG, ROELOF ARY IANS
 ; APPLICANT: KOERMAN, BERTUS PIETER
 ; APPLICANT: HOEKEMA, ANDREAS
 ; APPLICANT: VAN DER LAAN, JAN METSKE
 ; APPLICANT: VERWEIJ, JAN
 ; APPLICANT: DE VROOM, ERIK
 TITLE OF INVENTION: PROCESS FOR THE EFFICIENT PRODUCTION OF
 TITLE OF INVENTION: 7-ADCA VIA 3-(CARBOXYETHYLTHIO) PROPIONYL-7-ADCA
 NUMBER OF SEQUENCES: 15
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORRISON & FOERSTER
 STREET: 2000 PENNSYLVANIA AVENUE, NW
 CITY: WASHINGTON
 STATE: DC
 COUNTRY: USA
 ZIP: 20006-1888
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/591,501
 FILING DATE: 13-MAY-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: ADLER, REID G.
 REGISTRATION NUMBER: 30,988

REFERENCE/DOCKET NUMBER: 24615-20065.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 822-0168
TELEX: 90-4030 MRSNFORSMH
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 102 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-591-501-13

Alignment Scores:
Pred. No.: 1,22e-09 Length: 102
Score: 144.00 Matches: 27
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.80% Indels: 0
DB: 1 Gaps: 0

10719236-1_232-1164 (1-311) x US-08-591-501-13 (1-102)

Qy 284 Aspvalseleuaspiglythralthrphglnasprrp1eglyglyasentyrval 303
Db 1 GATGTCAGCCTGAGCGGAGAGCCGCCACGTTCCAGATTGATCGGGGCAACTACGTG 60

Qy 304 Aenlleargargthrserys 310
Db 61 AACATCCGCCGACATCCAG 81

RESULT 15
US-08-379-556A-1
Sequence 1, Application US/08379556A
Patent No. 5859329
GENERAL INFORMATION:
APPLICANT: HOLTON, TIMOTHY A.
APPLICANT: KEAM, LISA A.
TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOL
TITLE OF INVENTION: SYNTHASE ENZYMES AND USES THEREFORE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 GARDEN CITY PLAZA
CITY: GARDEN CITY
STATE: NEW YORK
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/379,556A
FILING DATE: 22-MAR-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: DIGIGLIO, FRANK S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 9592
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516)742-4343
TELEFAX: (516)742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1211 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO

FEATURE:
NAME/KEY: CDS
LOCATION: 59..1101
US-08-379-556A-1

Alignment Scores:
Pred. No.: 8,36e-08 Length: 1211
Score: 143.00 Matches: 59
Percent Similarity: 38.64% Conservative: 43
Best Local Similarity: 22.35% Mismatches: 114
Query Match: 8.74% Indels: 48
DB: 2 Gaps: 9

10719236-1_232-1164 (1-311) x US-08-379-556A-1 (1-1211)

Qy 27 Argasplysglyleuphetrleuthraspcysglyleuthraspthrghleuysser 46
Db 287 AAAGAGTGGGGATTTCCAACTGATCAACATGGCATTCCTGATGAGGCTATCGCGAT 346

Qy 47 Alalysaspllevalleasppheglnthrseryserglyualaglnlyalargalaval 66
Db 347 TTACGAAAGTAGGAAAGAGTTCTTGAACATGTTCCACAGAGAGAAAGAGCTGATT 406

Qy 67 Thr--Serprovalprothmetargarglyphe--Thrglyleuglusergluser 84
Db 407 GCCAAGACTCCAGATCAAGACACTTGAAGGCTATGAACTTCTTCAGAGAGAGTG 466

Qy 85 Thrhaglnlethrashnthrglysertryseraspyrsermetcysrtrysermetgly 104
Db 467 GAAGGCAG--AAAGGT 481

Qy 105 Thrhalaapnleuphe-----Proserglyasppheglnhrglletp 119
Db 482 TGGGTGATCATTTGTTCCATTAAGATTGGCTCTGCTGCGTCACTATCGTATTGG 541

Qy 120 -----Thrglntrypheaspparglntyrthr 128
Db 542 CCTAAAAACCTCTTCATACAGGAGAAACAGAAATATGAAAGAGATCGAGAA 601

Qy 129 Alaserargalavalalarggluvalleuarglathrgllythrghluproaspigly 148
Db 602 GTTGAGACAGAAATTTTAAAGACTTGTCTGGGCTTGGGCTTGAAGGCATGAATG 661

Qy 149 ValglualpheleuaspCysglu-----Proleuenuargpheargtryrphroglu 166
Db 662 ATAGAGCGAGCTGTGTGATGATGATTACTTGTGAAGATCAACTATTATCCACCA 721

Qy 167 ValprogluhtsargseralaglgluInProleuargmetalaProhetyraspleu 186
Db 722 TGCCCAAG-----CCGATTTGGCTCTGTGTGTGGCCCATACGACATG 769

Qy 187 Sermetvalthrleuileglnthrpocysalaennglyphevalserleuglnala 206
Db 770 TCATATATCACCAATCTT-----GTCCAAATGAAGTCCAGAGCTCCAAAGTG 817

Qy 207 GluvalglylyalaphethraspleuProtyrargProasplavalleuvalpheys 226
Db 818 TTCAAGGATGGCATTTGTATGATGTCAAGTACATACCAATGCTTAAATGTCCATATT 877

Qy 227 Glyalalalealathrleuvalthrghlyglnvallysalaproarghshlevalala 246
Db 878 GGTACCAAGTGTGATTTCTTACGAAATGCAAAATACAGAGTGTATACATAGACAAAG 937

Qy 247 Alaproargargaspllealaglyserserhrgthrserservalphepheleuarg 266
Db 938 GTGAACAAG-----GACAAAGCAAGATGTATGATGCGGTTTCTTGAG 982

Qy 267 Proasplaaap 270
Db 983 CCCCCGTACAG 994

Search completed: September 19, 2005, 18:02:59

Job time : 214 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: September 19, 2005, 16:58:00 ; Search time 741 Seconds
(without alignments)
2796.302 Million cell updates/sec

Title: 10719236-1_232-1164
Perfect score: 1636
Sequence: 1 MDTVPFTSLAEIQQGLHOD.....ATFQDWIGNMYNIRRTSKA 311

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 7389322 seqs, 331285599 residues
Total number of hits satisfying chosen parameters: 14778644

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

MODE=frame.p2n.model -DEV=xlh
-c=/cgn2_1/USPTO_spool_h/10719236-1/RRANS/runat_19092005_153200_7832/app_query.fasta_1.45
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blosum62
-RRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=spc -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000
-USER=10719236-1/RRANS.@CGN_1.1.480 @runat_19092005_153200_7832 -NCPU=6 -ICPU=3
-NO WMAP -LARGESUBJECT -NEG SCORES=0 -WAIT -DSEBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10J_NEW_PUB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1636	100.0	1230	16	US-10-105-319-1 Sequence 1, Appl1
2	1636	100.0	1230	18	US-10-719-237-1 Sequence 1, Appl1
3	1636	100.0	1230	18	US-10-719-236-1 Sequence 1, Appl1
4	1636	100.0	1230	19	US-10-719-238-1 Sequence 1, Appl1
5	1636	100.0	1230	20	US-10-789-071-1 Sequence 1, Appl1
6	169	100.3	1101	17	US-10-369-493-35926 Sequence 35926, A
7	168	10.3	29870	21	US-10-494-495-1 Sequence 1, Appl1
8	156	9.5	1335	18	US-10-425-114-22510 Sequence 22510, A
9	156	9.5	1559	20	US-10-425-115-35660 Sequence 35660, A
10	155	9.5	1256	18	US-10-425-114-16150 Sequence 16150, A
11	154.5	9.4	1281	19	US-10-437-963-43792 Sequence 43792, A
12	154	9.4	1346	18	US-10-425-114-1868 Sequence 1868, Ap
13	154	9.4	1346	18	US-10-425-114-20837 Sequence 20837, A
14	154	9.4	1376	18	US-10-425-114-22268 Sequence 22268, A
15	154	9.4	1430	20	US-10-425-115-35659 Sequence 35659, A
16	153	9.4	1294	19	US-10-437-963-34215 Sequence 34215, A
17	148.5	9.1	981	17	US-10-369-493-37730 Sequence 37730, A
18	148.5	9.1	1545	20	US-10-425-115-131728 Sequence 131728, A
19	146.5	9.0	1361	19	US-10-437-963-40955 Sequence 40955, A
20	146.5	9.0	1425	18	US-10-424-599-115110 Sequence 115110, A
21	145	8.9	1453	19	US-10-437-963-78719 Sequence 78719, A
22	144.5	8.8	1447	19	US-10-437-963-46106 Sequence 46106, A
23	143.5	8.8	1344	19	US-10-767-701-13412 Sequence 13412, A
24	143	8.7	1140	15	US-10-136-444-7 Sequence 7, Appl1
25	140.5	8.6	1287	19	US-10-437-963-74437 Sequence 74437, A
26	139.5	8.5	1387	18	US-10-425-114-25033 Sequence 25033, A
27	139.5	8.5	1407	20	US-10-431-115-162662 Sequence 162662, A
28	139.5	8.5	1449	17	US-10-431-973-44 Sequence 44, Appl1
29	135.5	8.3	777	17	US-10-369-493-41626 Sequence 41626, A
30	135.5	8.3	1059	19	US-10-437-963-93028 Sequence 93028, A
31	134.5	8.2	1322	19	US-10-437-963-38404 Sequence 38404, A
32	133.5	8.2	1293	22	US-10-819-773-4 Sequence 4, Appl1
33	133.5	8.2	1345	16	US-10-171-174A-13 Sequence 13, Appl1
34	133.5	8.2	1392	18	US-10-424-599-48151 Sequence 48151, A
35	132	8.1	806	17	US-10-369-493-42893 Sequence 42893, A
36	131.5	8.0	927	17	US-10-369-493-31336 Sequence 31336, A
37	131	8.0	1613	18	US-10-425-114-27030 Sequence 27030, A
38	131	8.0	2666	20	US-10-739-930-1684 Sequence 1684, Ap
39	130.5	8.0	1165	18	US-10-425-114-35376 Sequence 35376, A
40	130.5	8.0	1332	18	US-10-425-114-27849 Sequence 27849, A
41	130.5	8.0	1332	20	US-10-425-115-82503 Sequence 82503, A
42	130.5	8.0	1375	18	US-10-424-599-98643 Sequence 98643, A
43	130	7.9	3914	19	US-10-437-963-9728 Sequence 9728, Ap
44	129	7.9	960	17	US-10-369-493-37707 Sequence 37707, A
45	128.5	7.9	1075	19	US-10-437-963-40100 Sequence 40100, A

ALIGNMENTS

RESULT 1
US-10-105-319-1
; Sequence 1, Application US/10105319
; Publication No. US20030190695A1
; GENERAL INFORMATION:
; APPLICANT: Yunn-Bor Yang
; APPLICANT: Chia-Li Wei
; APPLICANT: Jyh-Shing Hsu
; APPLICANT: Yang-Chien Teai
; TITLE OF INVENTION: Mutated Penicillin Expandase and Process for Preparing 7-ADCA Usi
; TITLE OF INVENTION: the Same
; FILE REFERENCE: 6653-017-999
; CURRENT APPLICATION NUMBER: US/10/105,319
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 1

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Streptomyces clavuligerus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (232)...(1164)
US-10-105-319-1

Alignment Scores:
Pred. No.: 1,896-209 Length: 1230
Score: 1636.00 Matches: 311
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0

10719236-1_232-1164 (1-311) x US-10-105-319-1 (1-1230)

QY 1 MetAspThrThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAsp 20
DB 232 ATGGAACAGACGGTGGCCACCTTCAGCTGGCCGAATCCAGAGGGCTTGACACGAGAC 291
QY 21 GluPheArgArgCysLeuArgAspLysGlyLeuPheTyrLeuThrAspCysGlyLeuThr 40
DB 292 GAGTTCGCCAGGTGTCTGAGGAGCAAGGGCTCTTCTATCTGACGGACTGCGGTCTGACC 351
QY 41 AspThrGluLeuLysSerAlaLysAspIleValIleAspPhePheGluHisGlySerGlu 60
DB 352 GACACCGAGCTGAAGTCCGACAGAGCAATGCTCATCGCTTCTTGAGCACGGACGACGAG 411
QY 61 AlaGluLysArgAlaValThrSerProValProThrMetLysArgGlyPheThrGlyLeu 80
DB 412 GCGAGAGACCGCGCGTCACTCGCCGCTCCACCATGCGCGGAGCTTACCGAGGCTG 471
QY 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys 100
DB 472 GAGTCGAGAGACACCGCCACATCAACATACCGGCACCTACTCCGACTATCTCGATGTC 531
QY 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTyrThr 120
DB 532 TACTCGAAGGGCACCGCGGACCACTCTTCCGTCGAGCTTCCGAGCGGATCTGGACC 591
QY 121 GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
DB 592 CAGTACTTCGACCGCCAGTACACCGCTCCGCGGGGTCCCGGAGGTCTCTCGGGCG 651
QY 141 ThrGlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArg 160
DB 652 ACCGGAGCCGAGCCCGGAGGGGTGAGGCTTCTCTGACTGAGACCGCTGCTGCGG 711
QY 161 PheArgTyrPheProGlnValProGluHisArgSerAlaGluGluGlnProLeuArgMet 180
DB 712 TTCGGCTACTTCCCGCAGGTCCCGAGACCGGCGGAGGAGCCCTCCGCGATG 771
QY 181 AlaProHisTyrAspLeuSerMetValThrLeuIleGlnGlnIleThrProCysAlaAsnGly 200
DB 772 GCGCGGACCTAGACTGTCTGATGCTACCTCTCATCCAGACACCTCTGGCCACAGGC 831
QY 201 PheValSerLeuGlnAlaGluValGlyValAlaPheThrAspLeuProTyrArgProAsp 220
DB 832 TTTCGTCAACCTCCAGGCGGAGGTGGCGGGCGTTCAACGACCTGCTTACGTCGGAC 891
QY 221 AlaValLeuValPheCysGlyAlaAlaIleAlaThrLeuValThrGlyGlyValValLysAla 240
DB 892 GCCGTCTCTCTCTGCGGGCGCATCCGACCTCGGTGACCGGCGGCGCAGTCAAGGCC 951
QY 241 ProArgHisHisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260
DB 952 CCGCGGACCACTGTGCGGGCCCGCGAGGAGCAAGATAGCGGGGAGCGGACCTTCC 1011
QY 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280
```

```

|||||
DB 1012 AGTGTGTTCTTCTCCGCTCCCAACGGGACTTACCTTCCGTCGCGGGCGGAG 1071
QY 281 CysGlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrpIleGly 300
DB 1072 TGGCGCTTCATGTCAACCTCTGACGGGAGACCGCCACGTTCCAGATGGATCGGGGC 1131
QY 301 AsnTyrValAsnIleArgArgThrSerLysAla 311
DB 1132 AACTAGTGAACATCGCGCCACATCCAGGCA 1164

RESULT 2
US-10-719-237-1
; Sequence 1, Application US/10719237
; Publication No. US20040077034A1
; GENERAL INFORMATION:
; APPLICANT: Yunn-Bor Yang
; APPLICANT: Chia-Li Wei
; APPLICANT: Jyh-Shing Hsu
; APPLICANT: Ying-Chieh Tsai
; TITLE OF INVENTION: Mutated Penicillin Expandase and Process for Preparing 7-ADCA Unit
; TITLE OF INVENTION: The Same
; FILE REFERENCE: 6653-017-999
; CURRENT APPLICATION NUMBER: US/10/719,237
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: US/10/105,319
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Streptomyces clavuligerus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (232)...(1164)
US-10-719-237-1

Alignment Scores:
Pred. No.: 1,896-209 Length: 1230
Score: 1636.00 Matches: 311
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 18 Gaps: 0

10719236-1_232-1164 (1-311) x US-10-719-237-1 (1-1230)

QY 1 MetAspThrThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAsp 20
DB 232 ATGGAACAGACGGTGGCCACCTTCAGCTGGCCGAATCCAGAGGGCTTGACACGAGAC 291
QY 21 GluPheArgArgCysLeuArgAspLysGlyLeuPheTyrLeuThrAspCysGlyLeuThr 40
DB 292 GAGTTCGCCAGGTGTCTGAGGAGCAAGGGCTCTTCTATCTGACGGACTGCGGTCTGACC 351
QY 41 AspThrGluLeuLysSerAlaLysAspIleValIleAspPhePheGluHisGlySerGlu 60
DB 352 GACACCGAGCTGAAGTCCGACAGAGCAATGCTCATCGCTTCTTGAGCACGGACGAG 411
QY 61 AlaGluLysArgAlaValThrSerProValProThrMetLysArgGlyPheThrGlyLeu 80
DB 412 GCGAGAGAGCGCCCGTCACTCGCCGCTCCACCATGCGCGGCTTCAACGGGCTG 471
QY 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTyrSerAspTyrSerMetCys 100
DB 472 GAGTCGAGAGACACCGCCAGTACCAATACCGGAGCTACTCCACTGATGTCG 531
QY 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTyrThr 120
DB 532 TACTCGAAGGGCACCGCGGACCACTTCTTCCGTCGAGCTTCCAGCGGATCTGGACC 591
QY 121 GlnTyrPheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
```



```
Db 592 CAGTACTTGACCGCCAGTACACCGGCTCCGCGGGGTGCGCCGGGAGGTCTCTGGGGGG 651
Qy 141 ThrGlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArg 160
Db 652 ACCGGGACCGACCGGACCGGCGGGGTGCGGCTTCTCTGACCTGCGAGCCGCTGTCGGG 711
Qy 161 PheArgTyrrPheProGlnValProGluHisArgSerAlaGluGlnProLeuArgMet 180
Db 712 TTCGGCTACTTCCCGCAGGTCCTCCGAGCACCGGAGGCGGAGGAGCCCTGCGGATG 771
Qy 181 AlaProHisTyrrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGly 200
Db 772 GCGCGGCATTAAGACCTGTGATGTGATCACTTCATCCAGACACACCCCTGCGCCAAAGCC 831
Qy 201 PheValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrrArgProAsp 220
Db 832 TTCGTCAAGCTCCAGGCTCCAGGTCGCGGCTTCAACGACCTGCTACCGCTCCGCGAC 891
Qy 221 AlaValIleuValPheCysGlyValAlaIleAlaThrLeuValThrGlyGlyValIleValAla 240
Db 892 GCGGCTCTGCTCTTCTGCGGCGCATCGGACCTGTGTGATCCGCGGCGCAAGGTC 951
Qy 241 ProArgHisHisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260
Db 952 CCCCGGCACATGTCGCGGCGCCCGCGAGAGACAGATAGCGGGGCGAGCGCGCACTCC 1011
Qy 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280
Db 1012 AGTGTGTTCTTCTCTCGCTCCCAACGCGGACTTCACCTTCTCGCTCCGCTGCGCGGAG 1071
Qy 281 CysGlyPheAspValSerLeuAspGlyGlyThrAlaThrPheGluAspTrpIleGlyGly 300
Db 1072 TCGCGCTTCAATGTCAGCTTGAGCGGAGACCGCCACCTTCCAGGATTGATCGGGGGC 1131
Qy 301 AsnTyrrValAsnIleArgArgThrSerIleValAla 311
Db 1132 AACTACGTGAACATCCGCGGACATCCAAAGGCA 1164

RESULT 3
US-10-719-236-1
; Sequence 1, Application US/10719236
; Publication No. US2004007035A1
; GENERAL INFORMATION:
; APPLICANT: Yunn-Bor Yang
; APPLICANT: Chia-Li Wei
; APPLICANT: Jyh-Shing Hsu
; APPLICANT: Ying-Chieh Tsai
; TITLE OF INVENTION: Mutated Penicillin Expandase and Process for Preparing 7-ADCA Usi
; FILE REFERENCE: 6653-017-999
; CURRENT APPLICATION NUMBER: US/10/719,236
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: US/10/105,319
; PRIOR FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Streptomyces clavuligerus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (232)..(1164)
; US-10-719-236-1

Alignment Scores:
Pred. No.: 1,89e-209 Length: 1230
Score: 1636.00 Matches: 311
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 18 Gaps: 0
```

```
10719236-1_232-1164 (1-311) x US-10-719-236-1 (1-1230)
Qy 1 MetAspThrThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAsp 20
Db 232 ATGGACAGACGAGTGCCCACTTCAAGCTTGGCCCAACTCCAGCAGGGGCTTCACACAGAGC 291
Qy 21 GluPheAspArgCysLeuArgAspIleGlyLeuPheTyrrLeuThrAspCysGlyLeuThr 40
Db 292 GAGTTCGCGAGGTGTCTGAGGAGACAGGGCTCTTCTATCTGACGAGACTGTGGTCTGACC 351
Qy 41 AspThrGluLeuLysSerAlaLysAspIleValIleAspPhePheGluHisGlySerGlu 60
Db 352 GACACCGAGCTGAGAGTGGCCCAAGSACATGCTCATTCGACTTCTTCAGAGCAGGAGAGAG 411
Qy 61 AlaGluLysArgAlaValThrSerProValProThrMetArgArgIlePheThrGlyLeu 80
Db 412 GCGGAGAAAGCCGCGCTACCTCCGCTCCCAATGCGCCGCTTCACCGGCTTCACCGGGCTG 471
Qy 81 GluSerGluSerThrAlaGlnIleThrAspThrGlySerTyrrSerAspTyrrSerMetCys 100
Db 472 GAGTGGAGACACCGCCAGATCAACATACCGGACGCTACTCTCGACTTCTCGATGTGC 531
Qy 101 TyrrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTrpThr 120
Db 532 TACTCGATGGGCACCGGGGACCACTCTTCCGTCGCGGTGACTTCGAGCGGATCTGAGACC 591
Qy 121 GlnTyrrPheAspArgGlnTyrrThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
Db 592 CAGTACTTGCACCGCCAGTACACCGCTCCCGCGCGGTCCGCGGAGGTCTCTCGGGGGC 651
Qy 141 ThrGlyThrGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeuArg 160
Db 652 ACCGGGACCGAGCCCGAGGCGGGGTGAGGCTTCTCGACTGAGACCCTGCTGCGG 711
Qy 161 PheArgTyrrPheProGlnValProGluHisArgSerAlaGluGlnProLeuArgMet 180
Db 712 TTCGGCTACTTCCCGCAGGTCCTCCGAGACCGGACCGCGGAGAGCCCTCGGAGTG 771
Qy 181 AlaProHisTyrrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGly 200
Db 772 GCGCGGCATTAAGACCTGTGATGTGATCACTTCATCCAGACACACCTGCGCCAAAGCC 831
Qy 201 PheValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrrArgProAsp 220
Db 832 TTCGTCAAGCTCCAGGCTCCAGGTCGCGGCGCTTCAAGGACCTGCTACCGTCCGAGC 891
Qy 221 AlaValIleuValPheCysGlyValAlaIleAlaThrLeuValThrGlyGlyValIleValAla 240
Db 892 GCGGCTCTGCTCTTCTGCGGCGCATTCGAGACCTGTGTGACCGGCGGCGGCTCAAGGCC 951
Qy 241 ProArgHisHisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260
Db 952 CCCCGGCACATGTCGCGGCGCCCGCGAGAGACAGATAGCGGGGCGAGCACTCC 1011
Qy 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280
Db 1012 AGTGTGTTCTTCTCTCGCTCCCAACGCGGACTTCACTTCTCGCTCCGCTGCGGAG 1071
Qy 281 CysGlyPheAspValSerLeuAspGlyGlyThrAlaThrPheGluAspTrpIleGlyGly 300
Db 1072 TCGCGCTTCAATGTCAGCTTGAGCGGAGACCGCCACCTTCCAGGATTGATCGGGGGC 1131
Qy 301 AsnTyrrValAsnIleArgArgThrSerIleValAla 311
Db 1132 AACTACGTGAACATCCGCGGACATCCAAAGGCA 1164

RESULT 4
US-10-719-238-1
; Sequence 1, Application US/10719238
; Publication No. US20040121425A1
; GENERAL INFORMATION:
; APPLICANT: Yunn-Bor Yang
```

```
APPLICANT: Chia-Ii Wei
APPLICANT: Jyh-Shing Hsu
APPLICANT: Ying-Chieh Tsai
TITLE OF INVENTION: Mutated Penicillin Expandase and Process for Preparing 7-ADCA Usi
FILE OF INVENTION: the Same
FILE REFERENCE: 6653-017-999
CURRENT APPLICATION NUMBER: US/10/719,238
CURRENT FILING DATE: 2003-11-21
PRIOR APPLICATION NUMBER: US/10/105,319
PRIOR FILING DATE: 2002-03-26
NUMBER OF SEQ ID NOS: 1
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1230
TYPE: DNA
ORGANISM: Streptomyces clavuligerus
FEATURE:
NAME/KEY: CDS
LOCATION: (232)...(1164)
US-10-719-238-1
```

Alignment Scores:

Pred. No.:	1,89e-209	Length:	1230
Score:	1636.00	Matches:	311
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	19	Gaps:	0

10719236-1_232-1164 (1-311) x US-10-719-238-1 (1-1230)

```
Qy 1 MetAspThrThrValProThrPheSerLeuAlaGluLeuGlnGlyLeuHisGlnAsp 20
Db 232 ATGACACGACGAGCGGCCACCTTCGCGCTGGCCGAATCCAGCAGGCGCTGCACACGAGAC 291
Qy 21 GluPheAArgCybLeuAArgAspLySgLyLeuPheTyLeuThrAspCySgLyLeuThr 40
Db 232 GAGTTCGCCAGGTGCTCTAGAGGACAAAGGCGCTCTTCTATCTGACGCGATCGCGGCTGACC 351
Qy 41 AspThrGluLeuLySerAlaLyAspIleValIleAspPhePheGluHisGlySerGlu 60
Db 352 GACACCGAGCTGAAGTCGCGCCCAAGACATCTCATCGACTTCTTCAGACACGCGACGCGAG 411
Qy 61 AlaGluLyAspAlaValIleThrSerProValProThrMetAArgGlyPheThrGlyLeu 80
Db 412 GCGAGAAAGCGCGCTGCTCACTCGCCGCTCCCACTACAGCGCGGCTTCAACGAGGCTG 471
Qy 81 GluSerGluSerThrAlaGlnIleThrAsnThrGlySerTySerAspTySerMetCys 100
Db 472 GAGTCGAGAGACCGCCGCAATGATCAACAATCCGCGCACTACTCGACTCATGATGTC 531
Qy 101 TyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPheGluArgIleTrpThr 120
Db 532 TACTCGATGCGGACCGCGGACCAACTCTTCCGCTCGGTCGCTTCTGACGCGATCTGGACC 591
Qy 121 GlnTyPheAspArgGlnTyThrAlaSerArgAlaValAlaArgGluValLeuArgAla 140
Db 592 CAGTACTTCGACCGCCAGTACCGGCTCCGCGGGGTGCGCGGAGAGTCTTGCGGCG 651
Qy 141 ThrGlyThrGluProAspGlyGlyValAlaGluAlaPheLeuAspCySgLyProLeuLeuArg 160
Db 652 ACCGGGACGAGCCGCGAGCGCGGCTTCAAGGCTTCTCTGACTGCGACCGCTGCTGCGG 711
Qy 161 PheArgTyPheProGlnValProGluHisArgSerAlaGluGluGlnProLeuArgMet 180
Db 712 TTCCTCTCTCTCCCGAGGTCCCGAGCACCGCAACGCGCGAGACACCCCTCGGAGATG 771
Qy 181 AlaProHisTyPheAspLeuSerMetValThrLeuIleGlnGlnIleThrProCysAlaAsnGly 200
Db 772 GCGCGGCACTAGACCTGTGATGGTCACTCCATCAGACACACCTCGCGCCACAGGC 831
Qy 201 PheValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyPheArgProAsp 220
```

```
Db 832 TTGTCAGCTCCAGGCGGAGGTGCGGCGCGCTTACGAGACTGCTGCTACCTCCGAG 891
Qy 221 AlaValLeuValPheCySgLyAlaIleAlaThrLeuValThrGlyGlnValIleValAla 240
Db 892 GCGTCTCTCTCTGCGCGGCGCATCGCACCTGTGACCGCGCGCCAGGTCAAGGCC 951
Qy 241 ProArgHisHisValAlaAlaProAArgArgAspGlnIleAlaGlySerSerArgThrSer 260
Db 952 CCGGAGCACATGTCGCGGCCCCCGAGGACCAATACCGGCGGACGCGACCTTC 1011
Qy 261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280
Db 1012 AGTGTTCTTCTCTCGCTCGCTCCCAACGCGACTTCATCTTCGCTCCGCTGCGCGCGAG 1071
Qy 281 CySgLyPheAspValSerLeuAArgGlyGluThrAlaThrPheGlnAspTrpIleGly 300
Db 1072 TCGCGCTTCATGTACCTGACCTCGAGCGGAGACCGCACCGTTCAGAGATTGATCGGAGC 1131
Qy 301 AsnTyPheAlaAsnIleAArgTyThrSerTyAla 311
Db 1132 AACTACGTGAACATCCGCGCCGACATCCAGGCA 1164
```

RESULT 5

US-10-789-071-1
Sequence 1, Application US/10789071
Publication No. US20040259191A1

GENERAL INFORMATION:

APPLICANT: Yun-Bor Yang

APPLICANT: Chia-Ii Wei

APPLICANT: Jyh-Shing Hsu

APPLICANT: Ying-Chieh Tsai

TITLE OF INVENTION: Mutated Penicillin Expandase and Process for Preparing 7-ADCA Usi

FILE REFERENCE: 6653-017-999

CURRENT APPLICATION NUMBER: US/10/789,071

CURRENT FILING DATE: 2004-02-27

PRIOR APPLICATION NUMBER: US/10/105,319

PRIOR FILING DATE: 2002-03-26

NUMBER OF SEQ ID NOS: 1

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1

LENGTH: 1230

TYPE: DNA

ORGANISM: Streptomyces clavuligerus

FEATURE:

NAME/KEY: CDS

LOCATION: (232)...(1164)

US-10-789-071-1

Alignment Scores:

Pred. No.:	1,89e-209	Length:	1230
Score:	1636.00	Matches:	311
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	20	Gaps:	0

10719236-1_232-1164 (1-311) x US-10-789-071-1 (1-1230)

```
Qy 1 MetAspThrThrValProThrPheSerLeuAlaGluLeuGlnGlyLeuHisGlnAsp 20
Db 232 ATGACACGACGAGCGGCCACCTTCAGCTGCGGAACTCCAGACGAGGCGCTGCACGAGAC 291
Qy 21 GluPheAArgCybLeuAArgAspLySgLyLeuPheTyLeuThrAspCySgLyLeuThr 40
Db 292 GAGTTCGCCAGGTGCTCTAGAGGACAAAGGCGCTCTTCTATCTGACGCGATCTGAGCC 351
Qy 41 AspThrGluLeuLySerAlaLyAspIleValIleAspPhePheGluHisGlySerGlu 60
Db 352 GACACCGAGCTGAAGTCGCGCCAAAGACATCTCATCGACTTCTTCAGACGCGGACGAG 411
Qy 61 AlaGluLyAspAlaValIleThrSerProValProThrMetAArgGlyPheThrGlyLeu 80
```

```
Db      412 GCGAGAACGCGCGCTCACTGCCCTGCCACCACATGCGCGGCTTACCGGCGCTG 471
Qy      81 GIUSERGIUSERThRAAGInIleThRAenThRGlySerTyrSerAspTyrSerMetCys 100
Db      472 GAGTGGAGAGACACGCGCCAGATCAACAAATACCGGACGTACTCCGACTACTCGATGTC 531
Qy      101 TyrSerMetGlyThRAAspAsnLeuPheProSerGlyAspPheGluArgIleTrpThr 120
Db      532 TACTGATGGGACACCGCGGACAACTCTCCGCTCCGGTGACTCTGACGGGATCTCGACC 591
Qy      121 GlnTyrPheAspArgGlnTyrThRAAspArgAlaValAlaArgGluValLeuArgAla 140
Db      592 CAGTACTTCGACCGGACAGTACACCGGCTCCCGCGGCTGCGCGGAGGTCTCTGCGGCG 651
Qy      141 ThrGlyThRGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuArg 160
Db      652 ACCGGAGACCGACCGGACGCGGCGGCTGAGGCGCTTCTGACTGCGAGCGCTGCTGCGG 711
Qy      161 PheArgTyrPheProGlnValProGluHisArgSerAlaGluGluProLeuArgMet 180
Db      712 TTCCGCTACTTCCCGCAGGTCCCGCAGCACCGCAGCGGACGAGGACGCCCTGCGGATG 771
Qy      181 AlaProHisTyrAspLeuSerMetValThrLeuIleGlnGlnThrProCysAlaAsnGly 200
Db      772 GCGCGGACACTACGACTGTGATGTCACTTCATTCACAGACACCTGCGGCCAACCGGC 831
Qy      201 PheValSerLeuGlnAlaGluValGlyGlyAlaPheThrAspLeuProTyrArgProAsp 220
Db      832 TTCGTACAGCTTCAGAGCCGAGGTCCGCGCGGCTTTCACGAGCTCCCTACCGCTCGGAC 891
Qy      221 AlaValLeuValPheCysGlyValAlaIleAlaThrLeuValThrGlyGlyValValysAla 240
Db      892 GCCGCTCTGCTTCTTCTGCGGCGCCCATCCGACCTCGGTACCGCGCGGACGATCAAGGC 951
Qy      241 ProArgHisValValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSer 260
Db      952 CCCCCGACACATGTGCGGCGCCCCCGCAGGAGCACAGATGCGGCGACGCGCACCTCC 1011
Qy      261 SerValPhePheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGlu 280
Db      1012 AGTGTGCTTCTTCTCCGTCACACGCGGACTTCACCTTCTCGTCCGCTGCGCGGCGAG 1071
Qy      281 CysGlyPheAspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrpIleGlyGly 300
Db      1072 TCCGCTTCGATGTCACTGACCTGAGCGGACGAGACCGCACCTTCCAGATTTGATCGGAGC 1131
Qy      301 AsnTyrValAsnIleArgArgThrSerIysAla 311
Db      1132 AACTACGTGAACATCCGCGGACATCCAGAGCA 1164

RESULT 6
US-10-369-493-35926
; Sequence 35926, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 35926
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Mesorhizobium loti
US-10-369-493-35926
```

```
Alignment Scores:
Pred. No.: 7,31e-12 Length: 1101
Score: 169.00 Matches: 85
Percent Similarity: 36.63% Conservation: 41
Best Local Similarity: 24.71% Mismatches: 158
Query Match: 10.33% Indels: 60
DB: 17 Gaps: 14

10719236-1_232-1164 (1-311) x US-10-369-493-35926 (1-1101)
Qy      1 MetAspThrThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGln--- 19
Db      82 ATGCCCAAGATAGTGGCTCTGCTCGATCTGAGCCGCTGTGAACAGGCGGCGCTCGGAAGC 141
Qy      20 -----AspGluPheArgArgCysLeuArgAspGlyGlyLeuPheTyrLeuThr 35
Db      142 CGGACATTTCTAGGCGGACTTCGTTCCGCTCGCGGCGACATCGGCTTTTCTATCTTCC 201
Qy      36 AspCysGlyLeuThrAspThrGluLeuLysSerAlaLysAspIleValIleAspPhePhe 55
Db      202 GGACATGGCATCTCTGGCGGAGATTCAGCGAGGTGCTCACCGCATCGCGCAGTCTTT 261
Qy      56 GluHisGlySerGluAlaGluLysArgAlaValThrSerProValProThrMetArgArg 75
Db      262 GCCTTG---CCGAGCGCGAACAAGCTGCGCATCGAATGTCAGTCTCTCGCATTTCCG 318
Qy      76 GlyPheThrGlyLeuGluSerGluSerThrAlaGlnIleThraenThrGlySerTyrSer 95
Db      319 GGTATACGCGCGCGCGCGGCGGAACTGACTAG-----GGCAGAGAG 360
Qy      96 AspTyrSerMetCysTyrSerMetGlyThrAlaAspAsnLeuPheProSerGlyAspPhe 115
Db      361 GACTGGCGGCAACGCTTGACATCGGCTGACGCGCAAGCAATCGCGCAAGCGCGGGA 420
Qy      116 GluArgIleTrpThrGln---TyrPheAspArgGlnTyrThrAlaSer----- 130
Db      421 ACTCCCGCTGGACCCGCGCTGCAAGACCAAGAGTGGCGCGCGCTCCCGGATCTC 480
Qy      131 -----ArgAlaValAlaArgGluValLeuArg 139
Db      481 AAACCGCGCTCTTGCCTGCGACAGACAGTGCACCCGCGGACCTCCGCTCTGAG 540
Qy      140 AlaThrGlyThRGluProAspGlyGlyValGluAlaPheLeuAspCysGluProLeuLeu 159
Db      541 GCCTTCCGCGAATGCTCTGATCAGCCCGAGGATCGTTC-----GATCCGATTC-- 588
Qy      160 ArgPheArgTyrPheProGlnValProGluHisArg---SerAlaGluGluGlnProLeu 178
Db      589 -----TACACGAGGAGCCCAACATGCAATGAAAGATCTGCTACCCCGC 636
Qy      179 Arg-----MetAlaProHisTyrAspLeuSerMetValThr 190
Db      637 CGTACACAACCGCGCGGACGACGAGGTGCGCGGACAAAGATGATGTTTCTTCCAC 696
Qy      191 LeuIleGlnGlnThrProCysAlaAsnGlyPheValSerLeuGlnAlaGluValGly 210
Db      697 CTGCTGCTTCAG-----GATGACATTAAGGCGCTGCGATGATGACGCGC 744
Qy      211 AlaPheThrAspLeuProTyrArgProAspAlaValLeuValPheCysGlyAlaIleAla 230
Db      745 AGCTGGGTGATGTGAGACCCATTCCTCCGACGCTCTCTGCAACATCGCGGAATTTGTC 804
Qy      231 ThrLeuValThrGlyGlyValLysAlaProArgHisValAlaAlaProArgArg 250
Db      805 GAACGTGCGCTCAACGCGCTACTGCGCGGACCGTGATCGTGCGTACGCCGCC-- 861
Qy      251 AspGlnIleAlaGlySerSerArgThrSerSerValPhePheLeuArgProAsnAlaAsp 270
Db      862 -----GCCGCGTGCAGCGCATTTCCGTGCGCTTCTTTCACCGCGCAGGCTGAT 912
Qy      271 PheThr-----PheSerValProLeuAlaArgGluCysGlyPheAspVal 285
```

```
Db      913 GCGACGATACCGCTGCTTCGAGGAACTGGCGCGCAGCGCGCGCGGCA 972
Qy      286 SerleuaspGlygluThrAlaThrPhegluAspTrpIleGlyValAsnIle 305
Db      973 AGC---GATCCGGAACAATCCGCTGTTTCGAT---GTCCGAACAATGCTCAAAAGC 1026
Qy      306 ArgArgThrSer 309
Db      1027 GCGCTCGCGCTCG 1038

RESULT 7
US-10-494-495-1
; Sequence 1, Application US/10494495
; Publication No. US2005079498A1
; GENERAL INFORMATION:
; APPLICANT: SmithKline Beecham plc
; APPLICANT: The Governors of the University of Alberta
; TITLE OF INVENTION: Polynucleotides and Polypeptides Involved in Clavulanic
; TITLE OF INVENTION: Acid Biosynthesis and Use Thereof
; FILE REFERENCE: P32913
; CURRENT APPLICATION NUMBER: US/10/494,495
; PRIOR FILING DATE: 2004-05-04
; PRIOR APPLICATION NUMBER: PCT/GB02/04989
; PRIOR FILING DATE: 2002-11-06
; PRIOR APPLICATION NUMBER: GB 0126756.6
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: GB 0128776.2
; 2001-11-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 29870
; TYPE: DNA
; ORGANISM: Streptomyces clavuligerus
US-10-494-495-1

Alignment Scores:
Pred. No.:      1.32e-09      Length:      29870
Score:          168.00      Matches:      77
Percent Similarity: 39.32%      Conservative: 50
Best Local Similarity: 23.84%      Mismatches: 150
Query Match:    10.27%      Indels:      46
                        Gaps:      14
                        21

10719236-1_232-1164 (1-311) x US-10-494-495-1 (1-29870)

Qy      5 ValProThrPheSerleuAlaGluLeu-----GlnGlnGlyLeuHis 18
Db      1307 GTTCCGACCATCGACATCTCGCGCTGTTCCGAAACGACCGCGCGCAAGAACCGCGTC 1366
Qy      19 GlnAspGluPheArgArgCysLeuArgAspIleGlyLeuPheTyrLeuThrAspCysGly 38
Db      1367 GCCAGAGATACACAGGGGCTCGCGCGCTCGGCTTCTTACGCCACGACACGAGGC 1426
Qy      39 LeuThrAspTrpGluLeuLysSerAlaLysAspIleValIleAspPheGlnHisGly 58
Db      1427 GTG-----GACGTCCAGACAGCTCCAGACCGGTGGGAAGGAGTTCCACGGGCCATG 1477
Qy      59 SerGluAlaGluLys-----ArgAlaValThrSerProValProThrmec 73
Db      1478 ACCGACCGAGAGAACGACGACTCGCATCCAGCTGTCACACCGGACACCCCGCAGTGG 1537
Qy      74 ArgArgGlyPhe-----ThrGlyLeuGluSerGluSerThrAlaGlnIleThr 89
Db      1538 CGCAACGGCTTACTACAGCGCGTCCCGGCGAGAAAGCGCTGAGTCTTCTGTACCTC 1597
Qy      90 AsnThrGlySerTyrSerAspTyrSerMetCysTyrSerMetGlyThr-----Ala 106
Db      1598 AACCCGAGACTTCGGCGAGACCAACCCGATG---ATCGCGCGGGGAGACCGCATGACAGAG 1654
Qy      107 AspAsnLeuPheProSerGlyAsp-----PheGluArgIleTrpThrGlnTyr 122
Db      1655 GTGAACCTTGCGCCGACGAGAGCGGACCCGCGCTTCCGGCGTTTCGAGGGGCTAC 1714
```

```
Qy      123 PheAspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArgAlaThrGly 142
Db      1715 TACCGGAGAGTGTGAAGACTTCTCACCGCTGTCAATGGGGGCTGGCGCTCGGG 1774
Qy      143 ThrGluProAspGlyGlyValGluAla-----PheLeuAspCysGlu 156
Db      1775 ---AGCGCGAGACACTTCTTCAGCGGGCGCTCGCGAGCAGGACCTCCGTGTCGTC 1831
Qy      157 ProLeuLeuArgPheArgTyrPheProGlnValProGluHisArgSerAlaGlu--- 175
Db      1832 TCCTGTATCCGATACCCGTATCTGAGAGATACCCCGGTGAAAGCGGGTCCGACGCG 1891
Qy      176 GlnProLeuArgMetAlaProHisTyrAspLeuSerMetValThrLeuIleGlnIleThr 195
Db      1892 CAGCTCTGAGCTTCGAGACGACCTTCGAGAGCTTCGATTCACCGTGGCTTTCAGACC 1951
Qy      196 ProCysAlaAsnGlyPheValSerLeuGlnAlaGluValGlyAlaPheThrAspLeu 215
Db      1952 CAGGTGCAGAAC-----CTCCAGGTGAGACGCGTCGACGCGCTGGCGGACATC 1999
Qy      216 ProTyrArgProAspAlaValLeuValPheCysGlyAlaIleAlaThrLeuValThrGly 225
Db      2000 CCGACGTCGAGAACGACTTCTGTCATCTGCTCACTGCGTACCTACATGCGCATGTCACGAAC 2059
Qy      236 GlyGlnValLysAlaProArgHisHisValAlaAlaProArgArgAspGlnIleAlaGly 255
Db      2060 GACTACTCCCGCGCGCCCAACACCGCGTC-----AAGTTCGGAAC 2101
Qy      256 SerSerArgThrSerSerValPhePheLeuArgProAsnAlaAspPheThrPheSerVal 275
Db      2102 GCGAGAGGCTGTCCCTGCGCTTCTTCAACGCGGCGACGAGCGGCTCATCGAG--- 2158
Qy      276 ProLeuAlaArgGluCysGlyPheAspValSerLeuAspGlyIleThrAlaThrPheGln 295
Db      2159 CCGTTCGTCCGAGAG---GGCGGACGAGAGAGTGAAGAACGAGCCCTGTCTACGGG 2215
Qy      296 AspTrpIle 298
Db      2216 GACTACTCTC 2224

RESULT 8
US-10-425-114-22510
; Sequence 22510, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 22510
; LENGTH: 1335
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB587-225-H12_FLI
US-10-425-114-22510

Alignment Scores:
Pred. No.:      5.49e-10      Length:      1335
Score:          156.00      Matches:      78
Percent Similarity: 38.02%      Conservative: 49
Best Local Similarity: 23.35%      Mismatches: 119
Query Match:    9.54%      Indels:      88
                        Gaps:      16
```

10719236-1_232-1164 (1-311) x US-10-425-114-22510 (1-1335)

```
OY 4 ThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHISGlnAspGluPheArg 23
Db 205 ACCTCAGCGAGCGGTCTCTCCCTCCATCGTGCACCTCTCTCCGCGCGGACGAGGTGGCG 264
OY 24 ArgCysLeu-----ArgAspLysGlyLeuPheTyrLeuThrAspCysGlyLeu 39
Db 265 CGCGCCATCTCGAGGCGCGGACAGAGATCGGCTTCTTCCAGGTGTGAACACAGCGCTC 324
OY 40 ThrAspThrGluLeuLysSerAlaLysAspIleValIleAspPheGlnHisGlySer 59
Db 325 TCCCTGAGAGCGATGAGACATGAGACGCTGTGCGACAGACTTCTCCG-----CTG 378
OY 60 GluAlaGluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGly 79
Db 379 CGGCGGAGACAGAGCC-----GGG 399
OY 80 LeuGluSerGluSerThrAlaGlnIleThr-----AsnThr 91
Db 400 CTGTACTCGAGAGACACGGGAGGCGGACCCGAGTCTACTCCAGACCATGTTCCAGACG 459
OY 92 GlySerTyrSerAspTyrSerMetCysTyrTyrSerMetGlyThrAlaAspAsnLeuPhePro 111
Db 460 GCGCGCGAGAACTACTGGCGGACGCTCCGCTCCGCTGGCTGCTCC-----TTCCCC 510
OY 112 Ser---GlyAspPheGluArgIleTrp-----ThrGlnTyrPhe 123
Db 511 GCCCTCGGAGACAGCGCGCGGCTTGGCCGACAGCCCGGAGGCTCCGCGAGGTGGTG 570
OY 124 AspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArg-----139
Db 571 GAGAGCTTCACGTCGACAGCGCGGCGCTGGGAGATGAGATCTCGCTGCTGTCGAG 630
OY 140 AlaThrGlyThrGluProAsp-----GlyGlyValGluAlaPhe 152
Db 631 GGCCTCGGCTCGCGCGGACTACTGAGAGGAGACATGACGAGGCGG-----678
OY 153 LeuAspCysGluProLeuLeuArgPheArgTyrPheProGlnValProGlnHisArgSer 172
Db 679 -----GACGTGCTCTCCACGTCACCACTACCCGCTCGGCGCGGACCCCAACGCTC 729
OY 173 AlaGluGlnProLeuArgMetAlaProHisTyrAspLeuSerMetValThrLeuIle 192
Db 730 ACG-----CTCGGCTCGCGCGGCTGTCAGCGACCTCTCACTTGTCTC 777
OY 193 GlnGlnThrProCysAlaAsnGlyPheValSerLeuGlnAlaGluValGlyGlyAlaPhe 212
Db 778 -----CTCCCGACGATGTCCTCGC-----798
OY 213 ThrAspLeuProTyrArg-----ProAspAlaValLeu 223
Db 799 CTCGAGGTGCGCTTACAGGGGCGCATCGATCAGAGGTGAGCGCTGTCGCGCGCTTCCTC 858
OY 224 ValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyGlnValIleAlaProArgHis 243
Db 859 GTCACTTCGAGTGCAGCTTGAAGTGTGACGAACGGGATTTGAAGAGCATTCAGACGAC 918
OY 244 HisValAlaIleAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSerValPhe 263
Db 919 CGGAGTGAACCAAC-----CTGGAGTGGGCGGACAGCAACAGTGGCCACG 963
OY 264 PheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCysGlyPhe 283
Db 964 TTCATCATGCGCCACACGAGACTGCTCATCGCG-----CCGCGCGCGAGTTCCTCAGCAG 1020
OY 284 AspValSerLeuAspGlyValThrAlaThrPheGlnAspTrp 297
Db 1021 GACAACCGCGCGCTGCTACCGCACCTCTACCTTCGCGGACTTC 1062
```

RESULT 9
US-10-425-115-35660

```
/ Sequence 35660, Application US/10425115
/ Publication No. US20040214272A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plasmids
/ FILE REFERENCE: 38-21(53222)B
/ CURRENT APPLICATION NUMBER: US/10/425,115
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 369326
/ SEQ ID NO 35660
/ LENGTH: 1559
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(1559)
/ OTHER INFORMATION: unsure at all n locations
/ FEATURE:
/ OTHER INFORMATION: Clone ID: MFT4577_132519C.1
/ US-10-425-115-35660

Alignment Scores:
Pred. No.: 6.9e-10 Length: 1559
Score: 156.00 Matches: 78
Percent Similarity: 38.02% Conservative: 49
Best Local Similarity: 23.35% Mismatches: 119
Query Match: 9.54% Indels: 88
DB: 20 Gaps: 16

10719236-1_232-1164 (1-311) x US-10-425-115-35660 (1-1559)
OY 4 ThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHISGlnAspGluPheArg 23
Db 288 ACCTCAGCGAGCGGTCTCTCCCTCCATCGTGCACCTCTCTCCGCGCGGACGAGGTGGCG 347
OY 24 ArgCysLeu-----ArgAspLysGlyLeuPheTyrLeuThrAspCysGlyLeu 39
Db 348 CGCGCCATCTCGAGGCGCGGACAGAGATCGGCTTCTTCCAGGTGTGAACACAGCGCTC 407
OY 40 ThrAspThrGluLeuLysSerAlaLysAspIleValIleAspPheGlnHisGlySer 59
Db 408 TCCCTGAGAGCGATGAGACATGAGACGCTGTGCGACAGGTCTTCG-----CTG 461
OY 60 GluAlaGluLysArgAlaValThrSerProValProThrMetArgArgGlyPheThrGly 79
Db 462 CGGCGGAGACAGAGCC-----GGG 482
OY 80 LeuGluSerGluSerThrAlaGlnIleThr-----AsnThr 91
Db 483 CTGTACTCGAGAGACAGCGGAGGCGGACCCGAGTCTACTCCAGACCATGTTCCAGACG 542
OY 92 GlySerTyrSerAspTyrSerMetCysTyrTyrSerMetGlyThrAlaAspAsnLeuPhePro 111
Db 543 GCGCGCGAGAACTACTGGCGGACGCTCCGCTCCGCTGGCTGCTCC-----TTCCCC 593
OY 112 Ser---GlyAspPheGluArgIleTrp-----ThrGlnTyrPhe 123
Db 594 GCGCGCGGACAGCGCGCGGCTGCGCGGACGAGCCCGGAGGCTCCGCGAGGTGGTG 653
OY 124 AspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArg-----139
Db 654 GAGAGCTTCACGTCGAGAGCGCGGCGCTGGGAGTGAAGATCTCGCGCTGCTGCGAG 713
OY 140 AlaThrGlyThrGluProAsp-----GlyGlyValGluAlaPhe 152
Db 714 GGCCTCGGCTCGCGCGGACTACTGAGAGGAGACATCAGCGCGG-----761
OY 153 LeuAspCysGluProLeuLeuArgPheArgTyrPheProGlnValProGlnHisArgSer 172
Db 153 -----
```

Db 762 -----GACGCTGCTCTCCAGATCAACCACTACCCGCTGCCCCGACCCCAAGGCC 812
 Qy 173 AAGlUGlUGlInProleuArGMeAlaProHsTyArSpLeuSerMetValThrLeuIle 192
 Db 813 ACG-----CTCGGCTGCTCGCGCCGACGTGACCCGCAACCTTCACCTTCCTC 860
 Qy 193 GlnGlnThrProCysAlaAenGlyPheValSerLeuGlnIleValGlyAlaPhe 212
 Db 861 -----CTCCCGACAGTGGTCTCTGSC----- 881
 Qy 213 ThrAspLeuProTyArg-----ProAspAlaValLeu 223
 Db 882 CTCGAGGCGGCTTACAGGGCGCATGATCAGGGGAGCTGTGCCCCGGGCTTCGTG 941
 Qy 224 ValPheCysGlyAlaIleAlaThrLeuValThrglyGlnValValAlaProArgHis 243
 Db 942 GTCAACTTCGGATGCCAGCTTGAGATGTGACGAACGGGATCTGAAGAGATCGACAC 1001
 Qy 244 HisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSerValPhe 263
 Db 1002 CGGGTATGACCAAC-----CTGGAGTGGCGCGGACACACGTCGCCACG 1046
 Qy 264 PheLeuArgProAsnAlaAspPheThrPheSerValProleuAlaArgGluCyGlyPhe 283
 Db 1047 TTCATCATGCCACACGACGTGCTCATCGAC---CCCGCGCGGAGTTCTCAGCGAC 1103
 Qy 284 AspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrp 297
 Db 1104 GACAACCGCGCTGCTACCGCACCTTCACCTTCGCCGACCTTC 1145

RESULT 10

US-10-425-114-16150
 ; Sequence 16150, Application US/10425114
 ; Publication No. US20040034888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E.
 ; APPLICANT: Tabaska, Jack E.
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(5313)B
 ; CURRENT APPLICATION NUMBER: US/10/425.114
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 16150
 ; LENGTH: 1256
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: LIB3062-030-F8_FLI
 US-10-425-114-16150

Alignment Scores:

Pred. No.: 6.84e-10 Length: 1256
 Score: 155.00 Matches: 78
 Percent Similarity: 38.02% Conservative: 49
 Best Local Similarity: 23.35% Mismatches: 119
 Query Match: 9.47% Indels: 88
 DB: 18 Gaps: 16

10719236-1_232-1164 (1-311) x US-10-425-114-16150 (1-1256)

Qy 4 ThrValProThrPheSerLeuAlaGluLeuGlnGlnGlyLeuHisGlnAspGluPheArg 23
 Db 123 ACCTCAGGAGGAGGCTCTCCCTGCGCATGTCGACCTTCCTCCGCGCGGAGAGGTGGC 182
 Qy 24 ArgCysLeu-----ArgAspGlyGlyLeuPheTyTrLeuThrAspCysGlyLeu 39
 Db 183 CGCGCATCTCTCGAGCGCGGACAGAGATCGCTTCTTCAGGTGTGAACCAACGCGCTC 242

Qy 40 ThrAspThrGluLeuArgSerAlaLysAspIleValIleAspPheGlnHisGlySer 59
 Db 243 TCCCTGAGGCGGTGACGACGATGAGACGAGTGTCCAGAGATTCTTCGCG-----CTG 296
 Qy 60 GluAlaGluLysArgAlaValThrSerProValProThreArgArgGlyPheThrGly 79
 Db 297 CCGCGGAGAGACAAAGCC-----GGG 317
 Qy 80 LeuGluSerGluSerThrAlaGlnIleThr-----AsnThr 91
 Db 318 CTCTACTCGAGACACGGGAGGCGACCCGAGTCTACTCCAGACCATGTTTCACACG 377
 Qy 92 GlySerTySerAspTyArgSerMetCysTyArgSerMetGlyThrAlaAspLeuPhePro 111
 Db 378 GGGCGGAGAAAGTCTGGCGGACCTGCTCCGCTGCTGCTCC-----TTCCCC 428
 Qy 112 Ser---GlyAspPheGlnArgIleTrp-----ThrGlnTyPhe 123
 Db 429 GCCGTGCGGACAGCGCCGCGGCTGCGCCGACAAAGCCCGGAGGCTCCGGAGGTGTG 488
 Qy 124 AspArgGlnTyThrAlaSerArgAlaValAlaArgGluValLeuArg----- 139
 Db 489 GAGAGGTTCACGGTGCAGACGCGGCGCTGGGAGTGAAGATCTGTGCGCTGCTGCGAG 548
 Qy 140 AlaThrGlyThrGluProAsp-----GlyGlyValGluAlaPhe 152
 Db 549 GGCCTGCGGCTCCGCGCCGACCTACCTGGAAGGAGCATCAGCGCGGC----- 596
 Qy 153 LeuAspCysGluProleuLeuArgPheArgTyArgPheProGlnValProGluHisArgSer 172
 Db 597 -----GACGTGCTCTCCAGCTCAACCACTACCTCCGCTGCGCCGACCCCAAGCC 647
 Qy 173 AAGlUGlUGlInProleuArgMetAlaProHisTyArSpLeuSerMetValThrLeuIle 192
 Db 648 ACG-----CTCGGCTGCTCGCGCCGACGTGACCGCAACCTTCCTCATCTTCGTC 695
 Qy 193 GlnGlnThrProCysAlaAenGlyPheValSerLeuGlnIleValGlyAlaPhe 212
 Db 696 -----CTCCCGACAGTGGTCTCTGSC----- 716
 Qy 213 ThrAspLeuProTyArg-----ProAspAlaValLeu 223
 Db 717 CTCGAGGTGCTCCACAGGGCGCATGATCAGGGTGAAGCTGTGCCCCGCGCTTCGTC 776
 Qy 224 ValPheCysGlyAlaIleAlaThrLeuValThrglyGlnValValAlaProArgHis 243
 Db 777 GTCAACTTCGGATGCCAGCTTGAGATGTGACGAACGGGATCTGAAGAGATCGACAC 836
 Qy 244 HisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSerValPhe 263
 Db 837 CGGGTATGACCAAC-----CTGGAGTGGCGCGGACACACGTCGCCACG 881
 Qy 264 PheLeuArgProAsnAlaAspPheThrPheSerValProleuAlaArgGluCyGlyPhe 283
 Db 882 TTCATCATGCCACACGACGTGCTCATCGAC---CCCGCGCGGAGTTCTTCACCGAC 938

RESULT 11

US-10-437-963-43792
 ; Sequence 43792, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping

Qy 284 AspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrp 297
 Db 939 GACAACCGCGCTGCTACCGCACCTTCACCTTCGGGACCTTC 980

Page 9

Db	924	AATGGCAGGTACAGAGCCGTGAGCAACCGGGCCGTGGCCAGCGGCAGCACAGAC-----	977
Qy	254	ALAGlySerSerArGthrSerSerValPhePheLeuArgProAsnAlaAspPheThrPhe	273
Db	978	-----CGCAGTGTCCGTCGTCACTTCTACGCGCGGGGTACAGAGCTGACACTC	1025
Qy	274	SerValProLeuAlaArgLncysGlyPheAspValSerLeuAspGlyIuThrAlaThr	293
Db	1026	GGC-----CCGCTGCGC-----GACCTCGTCCGACGCGGGAGAGCGCGCAGG	1067
Qy	294	PheGlnAspTrpIleGlyIleValAsnIleArgArgThrSerIys	310

```

; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

```

Arg 23

Leu 39

Ser 59

79
GLY
|||

Thx 91

Pro 111
111
111
111

○
 ○
 ○
 ○
 ○
 ○

Qy 112 Ser---GLYAspPheGluArgIleTrp-----ThrGlnTyrPhe 123
 Db 534 GCGCTGGCGGACAGCGCGCGGCTGGCGGACAGCCCGGAGCTCCGGAGGTGGTG 593
 Qy 124 AspArgGlnTyrThrAlaSerArgAlaValAlaIleArgValLeuArg----- 139
 Db 594 GAGAGCTTACAGGTGACAGCGCGGAGCTGGGATGAGATCTCGCTGCTGGAG 653
 Qy 140 AlaThrGlyThrGluProAsp-----GlyGlyValGluAlaPhe 152
 Db 654 GGCCTGGCTCCCGCCGCACTAGAGGAGGACATCAGCGGCGG----- 701
 Qy 153 LeuAspCysGluProLeuLeuArgPheArgTyrPheProGlnValProGlnIleArgSer 172
 Db 702 -----GACGTGCTCTCCAGCTCAACCACTACCGCGCTCCCGGACCGCAAGGCC 752
 Qy 173 AlaGluGlnProLeuArgMetAlaProHISThrAspLeuSerMetValThrLeuIle 192
 Db 753 ACG-----CTGGGCTGCGCGGCACTGTGACCGCAACCTCTCACTTGGCTC 800
 Qy 193 GlnGlnThrProCysAlaAsnGlyPheValSerLeuGlnAlaGluValGlyGlyAlaPhe 212
 Db 801 -----CTCCCGACAGTGTACTGGC----- 821
 Qy 213 ThrAspLeuProTyrArg-----ProAspAlaValLeu 223
 Db 822 CTCGAGGTGCGCTTACAGGGGCGACTGATCCGGGTGAGCTGGCGCGGCGCTTGGTC 881
 Qy 224 ValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyValValAlaProArgHIS 243
 Db 882 GTCAACTTCGGATGCGAGCTTGAGTTGTGACGAACGGAGCTTGAAGAGATGAGACAC 941
 Qy 244 HisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSerValPhe 263
 Db 942 CGGGTGATGACCAAC-----CTGGAGTGGCGCGGACACACGGTGGCCACG 986
 Qy 264 PheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCysGlyPhe 283
 Db 987 TTCATCATGCCACACGAGCTGCTCATCGGC---CCGCGCGCGAGTTCTTCAAGGAC 1043
 Qy 284 AspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrp 297
 Db 1044 GACAACCGCGCGTGTACCGGACCTCTACCTTGGCGGACTTC 1085

RESULT 13
 US-10-425-114-20837
 ; Sequence 20837, Application US/10425114
 ; Publication No. US20040034888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jindong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E
 ; APPLICANT: Tabaska, Jack E
 ; APPLICANT: Cao, Yongwei
 ; TITLE OR INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(5313)B
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 20837
 ; LENGTH: 1346
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: LIB3245-358-A7_FLI
 US-10-425-114-20837

Alignment Scores: 1.03e-09 Length: 1346
 Pred. No.: 154.00 Matches: 78

Percent Similarity: 37.72% Conservative: 48
 Best Local Similarity: 23.35% Mismatches: 120
 Query Match: 9.41% Gaps: 88
 DB: 18 16
 10719236-1_232-1164 (1-311) x US-10-425-114-20837 (1-1346)
 Qy 4 ThrValProThrPheSerLeuAlaGluGlnGlnGlyLeuHISGlnAspGluPheArg 23
 Db 228 ACCTCAGGACGGTCTCTCCGCACTGAGACTCTCCCTCGGCGGACAGGTGGC 287
 Qy 24 ArgCysLeu-----ArgAspGlySerLeuPheThrLeuThrAspCysGlyLeu 39
 Db 288 CGGCGCATCTCGAGGCGCGGACAGAGATCGCTTTCAGAGTGGTGAACACGCGCTC 347
 Qy 40 ThrAspThrGluLeuSerAlaValAspIleValIleAspPheGlnHISGlySer 59
 Db 348 TCCCTGAGGCGATGAGGACATGAGAGAGAGCTGTGCSAGAGATTCTTCCG-----CTG 401
 Qy 60 GluAlaGluValArgAlaValThrSerProValProThrMetArgArgGlyPheThrGly 79
 Db 402 CCGCGGAGGACAAAGCC-----GGG 422
 Qy 80 LeuGluSerGluSerThrAlaGlnIleThr-----AsnThr 91
 Db 423 CTGTACTCGGAGGACACGGGAGAGGCGACCGGATCTACTCAGACATGTTTCAGACAG 482
 Qy 92 GlySerTyrSerAspTyrSerMetCysTyrSerMetGlyThrAlaAspAsnLeuPhePro 111
 Db 483 GGGCGGAGAGTACTGCGCGGCACTGCTCCGCTGGCTGCTGCTC-----TTCGCC 533
 Qy 112 Ser---GLYAspPheGluArgIleTrp-----ThrGlnTyrPhe 123
 Db 534 GCGCTGGCGGACAGCGCGGCGGCTGGCGGACAGCCCGGAGGTGGCGGAGGTGGTG 593
 Qy 124 AspArgGlnTyrThrAlaSerArgAlaValAlaArgGluValLeuArg----- 139
 Db 594 GAGAGCTTACAGGTGACAGCGCGGAGCTGGGATGAGATCTCGCTGCTGGAG 653
 Qy 140 AlaThrGlyThrGluProAsp-----GlyGlyValGluAlaPhe 152
 Db 654 GGCCTGGCTCCCGCCGCACTAGAGGAGGACATCAGCGGCGG----- 701
 Qy 153 LeuAspCysGluProLeuLeuArgPheArgTyrPheProGlnValProGlnIleArgSer 172
 Db 702 -----GACGTGCTCTCCAGCTCAACCACTACCGCGCTCCCGGACCGCAAGGCC 752
 Qy 173 AlaGluGlnProLeuArgMetAlaProHISThrAspLeuSerMetValThrLeuIle 192
 Db 753 ACG-----CTGGGCTGCGCGGCACTGTGACCGCAACCTCTCACTTGGCTC 800
 Qy 193 GlnGlnThrProCysAlaAsnGlyPheValSerLeuGlnAlaGluValGlyGlyAlaPhe 212
 Db 801 -----CTCCCGACAGTGTACTGGC----- 821
 Qy 213 ThrAspLeuProTyrArg-----ProAspAlaValLeu 223
 Db 822 CTCGAGGTGCGCTTACAGGGGCGACTGATCCGGGTGAGCTGGCGCGGCGCTTGGTC 881
 Qy 224 ValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyValValAlaProArgHIS 243
 Db 882 GTCAACTTCGGATGCGAGCTTGAGTTGTGACGAACGGAGCTTGAAGAGATGAGACAC 941
 Qy 244 HisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSerValPhe 263
 Db 942 CGGGTGATGACCAAC-----CTGGAGTGGCGCGGACACACGGTGGCCACG 986
 Qy 264 PheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCysGlyPhe 283
 Db 987 TTCATCATGCCACACGAGCTGCTCATCGGC---CCGCGCGCGAGTTCTTCAAGGAC 1043
 Qy 284 AspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrp 297

Db 1044 GACAAACCCGCGGTCTACCGCACCTTCACCTTCGCGCACTTC 1085

RESULT 14

US-10-425-114-22268

/ Sequence 22268, Application US/10425114

/ Publication No. US20040034888A1

/ GENERAL INFORMATION:

/ APPLICANT: Liu, Jindong

/ APPLICANT: Zhou, Yihua

/ APPLICANT: Kovalic, David K.

/ APPLICANT: Screen, Steven E

/ APPLICANT: Tabaska, Jack E

/ APPLICANT: Cao, Yongwei

/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

/ FILE REFERENCE: 38-21(5313)B

/ CURRENT APPLICATION NUMBER: US/10/425,114

/ NUMBER OF SEQ ID NOS: 73128

/ SEQ ID NO 22268

/ LENGTH: 1376

/ TYPE: DNA

/ ORGANISM: Zea mays

/ FEATURE:

/ OTHER INFORMATION: Clone ID: LIB357-058-F2_F1

US-10-425-114-22268

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1,076-09	1376	154.00	78	48	120	88	16
Percent Similarity:		37.72%					
Best Local Similarity:		23.35%					
Query Match:		9.41%					

DB: 18 Gaps: 16

10719236-1_232-1164 (1-311) x US-10-425-114-22268 (1-1376)

QY 4 ThrValProThrPheSerLeuAlaGluLeuGlnGlnGluLeuHisGlnAspGluPheArg 23

DB 179 ACCTCAGGCAACGGTCTCCCTCCATCGGACCTCTCCCTCGGCGGCAAGAGCTGG 238

QY 24 ArgCysLeu-----ArgAspArgGlyLeuPheArgGlyLeuPheArgCysGlyLeu 39

DB 239 CGCGCATCTCGAGGCGGCGGCAAGAGATCGGCTTCTTCAGATGGTGAACACGCGCTC 298

QY 40 ThrAspThrGluLeuLysSerAlaLysAspIleValIleAspPheGlnHisGlySer 59

DB 299 TCCCTGAGCGCATGCGAGCATGAGACGATGTCGACAGATGTTCTCCG-----CTG 352

QY 60 GluAlaGluLysArgAlaValAlaThrSerProValProThrMetArgArgGlyPheThrGly 79

DB 353 CGGCGGAGACAAAGCC-----GGG 373

QY 80 LeuGluSerGluSerThrAlaGlnIleThr-----AsnThr 91

DB 374 CTGTACTGGAGAGACACGCGGAGGCGGACCGGATCTACTCCAGACACATGTTGACACG 433

QY 92 GlySerTyrSerAspTyrSerMetCysTyrSerMetGlyThrAlaAspValLeuPhePro 111

DB 434 GCGCGCGAAGACTATGCGCGCATGCTCCGCTCCGCTGCGCTGCTCC-----TTCCCC 484

QY 112 Ser---GlyAspPheGluArgGlyLeuThr-----ThrGlnTyrPhe 123

DB 485 GCGCTCGGCGACAGCGCGCGCTGCGCGGACCAAGCCCGGAGGCTCCGCGAGGTGGT 544

QY 124 AspArgGlnTyrThrAlaSerArgAlaValAlaAspGluValLeuArg-----139

DB 545 GAGAGGTTTACAGGTGACAGCGCGGCGCTCGGAGATGAGATCTCGCGCTGTCGAG 604

QY 140 AlaThrGlyThrGluProAsp-----GlyGlyValGluAlaPhe 152

DB 605 GCGCTCGGCTCGCGCGGACTACCTGAGAGGAGCATATCAGCGCGCG-----652

QY 153 LeuAspCysGluProLeuLeuArgPheArgTyrPheProGlnValProGlnHisArgSer 172

DB 653 -----GACGTGTCTCCACGCTCAACCTACCGCCCTGCGCCGACCCCAAGCGC 703

QY 173 AlaGluGlnGlnProLeuArgMetAlaProHisTyrAspLeuSerMetValThrLeuIle 192

DB 704 ACG-----CTGGCGCTGCGCGCGCATGTCGACCGGACACTCTCTCACTTGGTTC 751

QY 193 GlnGlnThrProCysAlaAsnGlyPheValSerLeuGlnAlaGluValGlyAlaPhe 212

DB 752 -----CTCCACGATGTACTCGC-----772

QY 213 ThrAspLeuProTyrArg-----ProAspAlaValLeu 223

DB 773 CTGAGGTGCTCTTACAGGCGCATGTCGCGGTGAGCTGTGCGCGCGCTTTCG 832

QY 224 ValPheCysGlyValAlaIleAlaThrLeuValThrGlyGlnValIleValAlaProArgHis 243

DB 833 GTCACTTCGAGATGCGACCTTGAAGTTGTGACGAAACGGATCTTGAAGACATCGAGCAC 892

QY 244 HisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerValPhe 263

DB 893 CGGTGATGACCAAC-----CTGGAGTGGCGCGGACAAACGCTGCGCACG 937

QY 264 PheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCysGlyPhe 283

DB 938 TTCATCATGCCCAACGACGATGCTCATCGGC-----CCGCGCGCGAGTTCTCAAGCAC 994

QY 284 AspValSerLeuAspArgGlyThrAlaThrPheGlnAspTyr 297

DB 995 GACAAACCCGCGGTCTACCGCACCTTCACCTTCGCGCACTTC 1036

RESULT 15

US-10-425-115-35659

/ Sequence 35659, Application US/10425115

/ Publication No. US20040214272A1

/ GENERAL INFORMATION:

/ APPLICANT: La Rosa, Thomas J.

/ APPLICANT: Kovalic, David K.

/ APPLICANT: Zhou, Yihua

/ APPLICANT: Cao, Yongwei

/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

/ FILE REFERENCE: 38-21(53222)B

/ CURRENT APPLICATION NUMBER: US/10/425,115

/ NUMBER OF SEQ ID NOS: 369326

/ SEQ ID NO 35659

/ LENGTH: 1430

/ TYPE: DNA

/ ORGANISM: Zea mays

/ FEATURE:

/ OTHER INFORMATION: Clone ID: MRT4577_132518C.1

US-10-425-115-35659

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1,136-09	1430	154.00	78	48	120	88	16
Percent Similarity:		37.72%					
Best Local Similarity:		23.35%					
Query Match:		9.41%					

DB: 20 Gaps: 16

10719236-1_232-1164 (1-311) x US-10-425-115-35659 (1-1430)

QY 4 ThrValProThrPheSerLeuAlaGluLeuGlnGlnGluLeuHisGlnAspGluPheArg 23

DB 208 ACCTCAGGCAACGGTCTCCCTCCATCGGACCTTCTCCCTCGGCGGCAAGAGCTGG 267

QY 24 ArgCysLeu-----ArgAspArgGlyLeuPheArgGlyLeuPheArgCysGlyLeu 39

DB 268 CGCGCATCTCGAGGCGGCGGACAGAGATGCGCTTCTTCCAGGTGAGAACCAAGCGGCTC 327

```
Qy 40 ThrAspThrGluLeuYsSerAlaIysAspIleValIleAspPheGluHisGlySer 59
Db 328 TCCTGAGGCGATGACAGACATGAGAGCGGTGCGAGGATCTTCCG-----CTG 381
Qy 60 GluAlaGluYsArgAlaValThrSerProValProThrMetArgArgGlyPheThrGly 79
Db 382 CCGCGGAGGACAAAGGCC-----GGG 402
Qy 80 LeuGluSerGluSerThrAlaGlnIleThr-----AsnThr 91
Db 403 CTGTACTCGAGGACACCGGAGGCGACCGGATCTACTCCACGACCATGTTGACACG 462
Qy 92 GlySerTyrSerAspTyrSerMetCysTyrSerMetGlyThrAlaAspAsnLeuPhePro 111
Db 463 GCGCGGAGAAAGTACTGCGCGGACTGCTCCGCTGCGCTGCTCC-----TTCCCC 513
Qy 112 Ser---GlyAspPheGluArgIleTrp-----ThrGlnTyrPhe 123
Db 514 GCCGTCCGCGCACAGCGCGCGCTGGCCCGACAAAGCCCGAGAGCTCCGAGAGTGATG 573
Qy 124 AspArgGlnTyrThrAlaSerArgAlaValAlaArgIleValLeuArg----- 139
Db 574 GAGAGTTTCACGTCGACAGCCGCGGCTGGGATGGAGATCTGCGCTGTCGACG 633
Qy 140 AlaThrGlyThrGluProAsp-----GlyGlyValGluAlaPhe 152
Db 634 GGCCTCGGCTCCCGCCCGCTTACCTGGAAGGAGCATCAGCGCGGC----- 681
Qy 153 LeuAspCysGluProLeuLeuArgPheArgTyrPheProGlnValProGluHisArgSer 172
Db 682 -----GACGTGCTCTCCAGTCAACCATACCGCCCTGCGCCGACCCACGCC 732
Qy 173 AlaGluGluGlnProLeuArgMetAlaProHisTyrAspLeuSerMetValThrLeuIle 192
Db 733 ACG-----CTCGGCTGCGCGCATGTGACCGCACTCTCACCCTGCTC 780
Qy 193 GlnGlnThrProCysAlaAsnGlyPheValSerLeuGlnAlaGluValGlyAlaPhe 212
Db 781 -----CTCCCGACATGTACTGCG----- 801
Qy 213 ThrAspLeuProTyrArg-----ProAspAlaValLeu 223
Db 802 CTCGAGGTCCGCTACAGGCGCGCATGATCCGGGTGAGCTGTGCCGCGCCTTCGTC 861
Qy 224 ValPheCysGlyAlaIleAlaThrLeuValThrGlyGlyGlnValIysAlaProArgHis 243
Db 862 GTCAACTTCGAGTCCACCTTGAGGTGTGACGAACGGGATCTTGAAAGCATCGAGCAC 921
Qy 244 HisValAlaAlaProArgArgAspGlnIleAlaGlySerSerArgThrSerSerValPhe 263
Db 922 CCGGTGATGACCAAC-----CTGGAGTGGCGCGGACAAACGTTGCCACG 966
Qy 264 PheLeuArgProAsnAlaAspPheThrPheSerValProLeuAlaArgGluCysGlyPhe 283
Db 967 TTCATCATGCCACACACGAGCTGCTCATTCGCG-----CCGCGCGCGAGTTCTCAGCGAC 1023
Qy 284 AspValSerLeuAspGlyGluThrAlaThrPheGlnAspTrp 297
Db 1024 GACAAACCGCGCGTGTACCGGACCTCTCACCCTTCGCGACCTTC 1065
```

Search completed: September 19, 2005, 21:13:51
Job time : 758 secs